

# **Document made available under the Patent Cooperation Treaty (PCT)**

International application number: PCT/EP04/013630

International filing date: 01 December 2004 (01.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: BE  
Number: PCT/BE03/00207  
Filing date: 01 December 2003 (01.12.2003)

Date of receipt at the International Bureau: 04 April 2005 (04.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

# ROYAUME DE BELGIQUE

## MINISTERE DES AFFAIRES ECONOMIQUES ADMINISTRATION DE LA POLITIQUE COMMERCIALE



EP/04/13630

Il est certifié que les annexes à la présente sont la copie fidèle de documents que détient l'Office de la Propriété Industrielle.

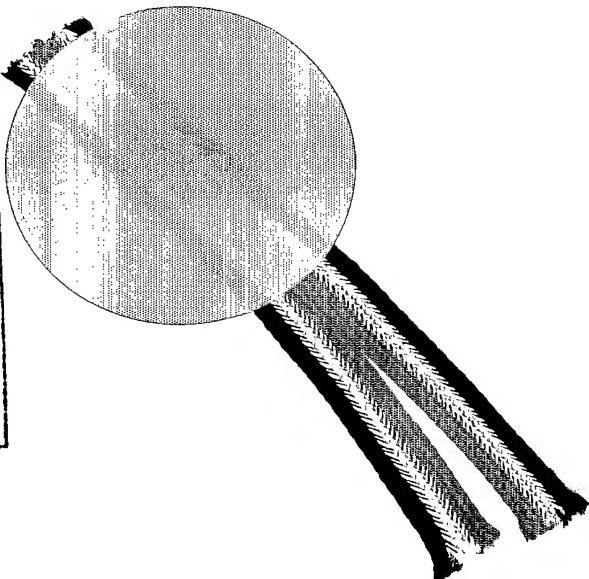
Bruxelles, le -7.-12-2004

Pour le Conseiller de l'Office  
de la Propriété industrielle

Le fonctionnaire délégué,

A handwritten signature in black ink, appearing to read "PETIT M." followed by a surname.

PETIT M.  
Conseiller adjoint



# COPY FOR THE RECEIVING OFFICE

1/6

## PCT REQUEST

AIC-024-PCT

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

0	For receiving Office use only	
0-1	International Application No.	P C T / B E 0 3 / 0 0 2 0 7
0-2	International Filing Date	01 DEC. 2003 (01-12-2003)
0-3	Name of receiving Office and "PCT International Application"	RO/BE-PCT INTERNATIONAL APPLICATION
0-4	Form - PCT/RO/101 PCT Request	
0-4-1	Prepared using	PCT-EASY Version 2.92 (updated 01.07.2003)
0-5	Petition	The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty
0-6	Receiving Office (specified by the applicant)	Intellectual Property Office (Belgium) (RO/BE)
0-7	Applicant's or agent's file reference	AIC-024-PCT
I	Title of invention	A HIGHLY OPTIMIZED METHOD FOR MODELLING, I.A. ANALYSIS AND SYNTHESIS, A WINDOWED SIGNAL WITH A VARIABLE LENGTH, AND THE USE OF THIS METHOD FOR PITCH ESTIMATION, SOURCE SEPARATION, AUDIO EFFECTS, AUDIO CODING, AUDIO ANNOTATION AND TRANSCRIPTION
II	Applicant	
II-1	This person is:	applicant only
II-2	Applicant for	all designated States except US
II-4	Name	AIC
II-5	Address:	Drie Eikenstraat 661 B-2650 Edegem Belgium
II-6	State of nationality	BE
II-7	State of residence	BE
II-8	Telephone No.	+32 3 826 93 04
II-9	Facsimile No.	+32 3 826 93 78
II-10	e-mail	info@antwerpinnovation.com

## PCT REQUEST

2/6

AIC-024-PCT

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

III-1	Applicant and/or inventor	
III-1-1	This person is:	
III-1-2	Applicant for	
III-1-4	Name (LAST, First)	
III-1-5	Address:	
III-1-6	State of nationality	
III-1-7	State of residence	
IV-1	Agent or common representative; or address for correspondence The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:	
IV-1-1	Name (LAST, First)	
IV-1-2	Address:	
IV-1-3	Telephone No.	
IV-1-4	Facsimile No.	
IV-1-5	e-mail	
IV-2	Additional agent(s)	
IV-2-1	Name(s)	
V	Designation of States	
V-1	Regional Patent (other kinds of protection or treatment, if any, are specified between parentheses after the designation(s) concerned)	

applicant and inventor  
US only  
D'HAES, Wim  
Maastrichtersteenweg 223/2  
B-3500 Hasselt  
Belgium  
BE  
BE

agent  
BRANTS, Johan, Philippe, Emile  
De Clercq, Brants & Partners  
E. Gevaertdreef 10a  
B-9830 Sint-Martens-Latem  
Belgium  
+32 9 280 23 40  
+32 9 280 23 45  
[info@dcbpatent.com](mailto:info@dcbpatent.com)

additional agent(s) with same address as first named agent  
DE CLERcq, Ann; VIAENE, Ann

AP: GH GM KE LS MW MZ SD SL SZ TZ UG ZM  
ZW and any other State which is a Contracting State of the Harare Protocol and of the PCT  
EA: AM AZ BY KG KZ MD RU TJ TM and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT  
EP: AT BE BG CH&LI CY CZ DE DK EE ES FI  
FR GB GR HU IE IT LU MC NL PT RO SE SI  
SK TR and any other State which is a Contracting State of the European Patent Convention and of the PCT  
OA: BF BJ CF CG CI CM GA GN GQ GW ML MR  
NE SN TD TG and any other State which is a member State of OAPI and a Contracting State of the PCT

## PCT REQUEST

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

V-2	National Patent (other kinds of protection or treatment, if any, are specified between parentheses after the designation(s) concerned)	AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH&LI CN CO CR CU CZ DE DK DM DZ EC EE EG ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NI NO NZ OM PG PH PL PT RO RU SC SD SE SG SK SL SY TJ TM TN TR TT TZ UA UG US UZ VC VN YU ZA ZM ZW
V-5	Precautionary Designation Statement  In addition to the designations made under items V-1, V-2 and V-3, the applicant also makes under Rule 4.9(b) all designations which would be permitted under the PCT except any designation(s) of the State(s) indicated under item V-6 below. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit.	
V-6	Exclusion(s) from precautionary designations	NONE
VI	Priority claim	NONE
VII-1	International Searching Authority Chosen	European Patent Office (EPO) (ISA/EP)
VIII	Declarations	Number of declarations
VIII-1	Declaration as to the identity of the inventor	-
VIII-2	Declaration as to the applicant's entitlement, as at the international filing date, to apply for and be granted a patent	-
VIII-3	Declaration as to the applicant's entitlement, as at the international filing date, to claim the priority of the earlier application	-
VIII-4	Declaration of inventorship (only for the purposes of the designation of the United States of America)	1
VIII-5	Declaration as to non-prejudicial disclosures or exceptions to lack of novelty	-

## PCT REQUEST

4/6

AIC-024-PCT

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

VIII-4-1	<p>Declaration: Inventorship (only for the purposes of the designation of the United States of America)</p> <p>Declaration of inventorship (Rules 4.17(iv) and 51bis.1(a)(iv)) for the purposes of the designation of the United States of America:</p> <p>I hereby declare that I believe I am the original, first and sole (if only one inventor is listed below) or joint (if more than one inventor is listed below) inventor of the subject matter which is claimed and for which a patent is sought.</p> <p>This declaration is directed to the international application of which it forms a part (if filing declaration with application).</p> <p>I hereby declare that my residence, mailing address, and citizenship are as stated next to my name.</p> <p>I hereby state that I have reviewed and understand the contents of the above-identified international application, including the claims of said application. I have identified in the request of said application, in compliance with PCT Rule 4.10, any claim to foreign priority, and I have identified below, under the heading "Prior Applications," by application number, country or Member of the World Trade Organization, day, month and year of filing, any application for a patent or inventor's certificate filed in a country other than the United States of America, including any PCT international application designating at least one country other than the United States of America, having a filing date before that of the application on which foreign priority is claimed.</p>
VIII-4-1 -1	Prior applications:

## PCT REQUEST

AIC-024-PCT

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

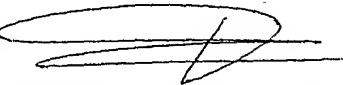
	<p>I hereby acknowledge the duty to disclose information that is known by me to be material to patentability as defined by 37 C.F.R. § 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the PCT international filing date of the continuation-in-part application.</p> <p>I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.</p>	
VIII-4-1 -1-1	Name:	
VIII-4-1 -1-2	Residence: (city and either US State, if applicable, or country)	
VIII-4-1 -1-3	Mailing address:	
VIII-4-1 -1-4	Citizenship:	
VIII-4-1 -1-5	Inventor's Signature: (if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)	
VIII-4-1 -1-6	Date: (of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)	

6/6

## PCT REQUEST

AIC-024-PCT

Original (for SUBMISSION) - printed on 01.12.2003 06:28:04 PM

IX	Check list	number of sheets	electronic file(s) attached
IX-1	Request (including declaration sheets)	6	-
IX-2	Description	36	-
IX-3	Claims	4	-
IX-4	Abstract	1	EZABST00.TXT
IX-5	Drawings	13	-
IX-7	TOTAL	60	
	Accompanying items	paper document(s) attached	electronic file(s) attached
IX-8	Fee calculation sheet	✓	-
IX-17	PCT-EASY diskette	-	Diskette
IX-19	Figure of the drawings which should accompany the abstract		
IX-20	Language of filing of the international application	English	
X-1	Signature of applicant, agent or common representative		
X-1-1	Name (LAST, First)	BRANTS, Johan, Philippe, Emile	

## FOR RECEIVING OFFICE USE ONLY

10-1	Date of actual receipt of the purported international application	01 DEC. 2003 (01-12-2003)
10-2	Drawings:	
10-2-1	Received	Received
10-2-2	Not received	
10-3	Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application	
10-4	Date of timely receipt of the required corrections under PCT Article 11(2)	
10-5	International Searching Authority	ISA/EP
10-6	Transmittal of search copy delayed until search fee is paid	

## FOR INTERNATIONAL BUREAU USE ONLY

11-1	Date of receipt of the record copy by the International Bureau	
------	--	--

A HIGHLY OPTIMIZED METHOD  
 FOR MODELLING, I.A. ANALYSIS AND SYNTHESIS,  
 A WINDOWED SIGNAL WITH A VARIABLE LENGTH,  
 AND THE USE OF THIS METHOD FOR PITCH ESTIMATION,  
 SOURCE SEPARATION, AUDIO EFFECTS, AUDIO CODING,  
 AUDIO ANNOTATION AND TRANSCRIPTION

FIELD OF THE INVENTION

The present invention relates to the modelling (analysis and synthesis) of musical signals and speech. The analysis method computes of amplitudes, phases and frequencies using a non linear least squares estimation technique. The synthesis comprises the reconstruction of the signal from these parameters. The method allows a variable window length and is highly optimized, reducing the time complexity of the basic least squares method from  $\mathcal{O}(N^3)$  to  $N \log(N)$ . The present invention further relates to computer programs and devices therefore.

BACKGROUND TO THE INVENTION

The sinusoidal modelling of sound signals such as music and speech is a powerful tool for parameterizing sound sources. Once a sound has been parameterized, it can be synthesized for example, with a different pitch and duration.

A sampled short time signal on which a window  $w_n$  is applied may be represented by a model  $\tilde{x}_n$  consisting of a sum of sinusoids which are characterized by their frequency  $\omega_k$ , phase  $\phi_k$  and amplitude  $a_k$ ,

$$\tilde{x}_n = w_n \sum_{k=0}^{K-1} a_k \cos\left(2\pi\omega_k \frac{n - n_0}{N} + \phi_k\right) + r_n \quad (1)$$

The offset value  $n_0$  allows the origin of the timescale to be placed exactly in the middle of the window. The noise residual is denoted  $r_n$ . For a signal with length  $N$ ,  $n_0$  equals  $\frac{N-1}{2}$ .

If the signal would be synthesized by a bank of oscillators, the complexity would be  $\mathcal{O}(NK)$  with  $N$  being the number of samples and  $K$  the number of sinusoidal components.

As described in patent WO 93/03478, the computational efficiency of the synthesis can be improved by using the inverse fourier transform. By using a window that has a small bandwidth in the frequency domain, the frequency response of each partial can be computed in constant time. A computation technique was disclosed in WO 93/03478 by using look-up tables.

Patent application WO 90/13887 discloses the estimation of the amplitudes by detecting individual peaks in the magnitude spectrum, and performing a parabolic interpolation to refine the frequency and amplitude values. In WO 93/04467 and WO 95/30983 a least means squares is presented for individual sinusoidal components which is applied iteratively, subtracting a single sinusoidal component in each iteration.

Presently, methods of the prior art achieve synthesis by inverse Fourier transformation using a fixed window length. However, such techniques are disadvantageous because "smearing" can occur, for example, when the length of the window is too large. This leads to an inaccurate synthesis of the sound.

Techniques of the prior art for analysing sound signals cannot resolve overlapping frequency responses of sinusoidal components with close frequencies. These overlapping peaks occur when multiple sound sources are mixed together or in monophonic signals with strong reverberation. Also when very small windows are used, all frequency responses can overlap even for a monophonic sound.

Methods of the prior art use a basic least squares computation for analysing the signal. This requires a time complexity  $\mathcal{O}(K^2N)$  and a space complexity  $\mathcal{O}(K^2)$ . Due to the complexity of the calculation, methods of the prior art make heavy demands on the computing power.

There is a need for a method for analysing and synthesising sound signals that overcomes these problems.

## SUMMARY OF THE INVENTION

The present invention relates to the modelling (analysis and synthesis) of musical signals and speech and provides therefore a method for modelling, i.a. analyzing and/or synthesizing, a windowed signal with a variable length, by computing the frequencies and complex amplitudes from the signal by using an optimized least squares method, whereby the method is executed using one or more variable length windows and has a complexity  $\mathcal{O}(N \log N)$ .

The numerous computational optimizations that are disclosed in this invention allow to reduce the original complexity of the non-linear least squares method which is  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N \log N)$ . It is known that the great advantage of the least squares method is that it is able to resolve overlapping peaks and is therefore more accurate than iterative methods. Today, iterative methods are widely used because of their smaller computational complexity. The present invention allows to speed up the superior least squares method to a computational complexity comparable with iterative methods.

The invention improves several applications such as accurate pitch estimation, parametric audio coding, source separation, audio effects and automated annotation and transcription

The method computes the amplitudes, phases and frequencies using a non linear least squares estimation technique which is depicted in Figure 14.

Two types of models are distinguished in the invention. The first model consists of a superposition of an arbitrary set of sinusoidal components. The second model consists of a set periodic sources that are described by a harmonic series of sines. All sines of a given source are a multiple of the fundamental frequency.

For the harmonic model, a multipitch estimator provides a set of frequencies corresponding to the fundamental frequency of each source. For the non-harmonic model, the initial frequency values are obtained by detecting local maxima in the sampled spectrum.

In a pre-processing step of the amplitude estimation, the frequencies are sorted and frequencies that occur multiple times are omitted. For each sinusoidal component the number of frequencies are counted that fall into its main lobe. The maximum of this number over all components yields that number of diagonal bands that are considered for the amplitude estimation.

The amplitude calculator computes the amplitudes of the sinusoids using a least squares method. We show that by incorporating a window with a band limited frequency response in the least squares derivation, a band diagonal system of equations is obtained. Recall that

the number of diagonal bands was determined during the pre-processing. This results in the fact that now the equations can be solved in linear time. The computation of the individual elements in the system is optimized using fast frequency response computation methods.

The estimated amplitudes and frequencies are then used to compute the spectrum of the signal model. The difference with the observed spectrum yields the spectrum of the residual signal which is used for the optimization of the frequencies.

The frequencies are optimized by making a local quadratic approximation of the error function and are iteratively adjusted using Newton's method. This method requires the computation of the gradient and Hessian matrix of the error function. For the computation of the elements of the Hessian matrix and gradient, fast frequency response calculators are used. The band limited property of the window results in a band diagonal Hessian matrix which can be inverted in linear time. The optimized frequencies are used in the next iteration. This is repeated until a stopping criterium is met.

## SUMMARY OF THE INVENTION

The present invention relates to the modelling (analysis and synthesis) of musical signals and speech and provides therefore a method for modelling, i.a. analyzing and/or synthesizing, a windowed signal with a variable length, by computing the frequencies and complex amplitudes from the signal by using an optimized least squares method, whereby the method is executed using one or more variable length windows and has a complexity  $\mathcal{O}(N \log N)$ .

The numerous computational optimizations that are disclosed in this invention allow to reduce the original complexity of the non-linear least squares method which is  $\mathcal{O}(N^3)$  to  $\mathcal{O}(N \log N)$ . It is known that the great advantage of the least squares method is that it is able to resolve overlapping peaks and is therefore more accurate than iterative methods. Today, iterative methods are widely used because of their smaller computational complexity. The present invention allows to speed up the superior least squares method to a computational complexity comparable with iterative methods.

The invention improves several applications such as accurate pitch estimation, parametric audio coding, source separation, audio effects and automated annotation and transcription

The method computes the amplitudes, phases and frequencies using a non linear least squares estimation technique which is depicted in Figure 14.

Two types of models are distinguished in the invention. The first model consists of a superposition of an arbitrary set of sinusoidal components. The second model consists of a set periodic sources that are described by a harmonic series of sines. All sines of a given source are a multiple of the fundamental frequency.

For the harmonic model, a multipitch estimator provides a set of frequencies corresponding to the fundamental frequency of each source. For the non-harmonic model, the initial frequency values are obtained by detecting local maxima in the sampled spectrum.

In a pre-processing step of the amplitude estimation, the frequencies are sorted and frequencies that occur multiple times are omitted. For each sinusoidal component the number of frequencies are counted that fall into its main lobe. The maximum of this number over all components yields that number of diagonal bands that are considered for the amplitude estimation.

The amplitude calculator computes the amplitudes of the sinusoids using a least squares method. We show that by incorporating a window with a band limited frequency response in the least squares derivation, a band diagonal system of equations is obtained. Recall that

the number of diagonal bands was determined during the pre-processing. This results in the fact that now the equations can be solved in linear time. The computation of the individual elements in the system is optimized using fast frequency response computation methods.

The estimated amplitudes and frequencies are then used to compute the spectrum of the signal model. The difference with the observed spectrum yields the spectrum of the residual signal which is used for the optimization of the frequencies.

The frequencies are optimized by making a local quadratic approximation of the error function and are iteratively adjusted using Newton's method. This method requires the computation of the gradient and Hessian matrix of the error function. For the computation of the elements of the Hessian matrix and gradient, fast frequency response calculators are used. The band limited property of the window results in a band diagonal Hessian matrix which can be inverted in linear time. The optimized frequencies are used in the next iteration. This is repeated until a stopping criterium is met.

## BRIEF SUMMARY OF THE FIGURES

Figure 1 depicts the frequency responses of the Blackmann- Harris window and the first and second derivative of frequency response.

Figure 2: depicts the frequency responses of the zero padded Blackmann- Harris window,  
5 the frequency response of the squared window and its second derivative.

Figure 3 depicts the theoretic motivation for the scaled look-up table.

Figure 4 depicts the scaled table look-up.

Figure 5 depicts the analytic frequency response computation.

Figure 6 depicts the sampled frequency response interpolation.

10 Figure 7 depicts the optimized synthesis method.

Figure 8 depicts the optimized amplitude computation.

Figure 9 depicts the pre-processing for the amplitude computation.

Figure 10 depicts the calculation of the initial frequency values.

Figure 11 depicts the frequency optimization for the non-harmonic model.

15 Figure 12 depicts the frequency optimization for the harmonic model.

Figure 13 depicts a subroutine of the frequency optimization for the harmonic model.

Figure 14 depicts the complete Analysis/Synthesis method.

Figure 15 depicts some applications for which the present invention will provide a considerable improvement. These applications are: 1) high accuracy pitch estimation, 2) audio  
20 coding, 3) audio effects and 4) source separation.

## DETAILED DESCRIPTION OF THE INVENTION

### 1 Synthesis

#### 1.1 Introduction

The present invention provides an extension to inverse fourier transform synthesis by allowing variable window lengths that not necessarily need to be a power of two. Since this is required for the inverse fast fourier transform, a window with length  $M$  is zero padded up to a length  $N = 2^{\lceil \log_2(M) \rceil}$ . This results however in a different frequency response for each frame. Three methods are disclosed for computing the frequency response of a window with a variable length. Any of these methods can be used to compute the frequency responses of the windows that are applied throughout the invention. In the illustrations they are denoted as *variable length frequency response calculators*.

#### 1.2 Frequency Response of the Window

Typically, a window is chosen that has a frequency response with a limited bandwidth containing the main lobe. The response outside this band is very small so that it can be neglected. This justifies the fact that only the main lobe of the frequency response must be computed in the synthesized spectrum  $\tilde{X}_m$ . In particular, but not exclusively, we cite the Blackmann-Harris window

$$w_n = a + b \cos\left(2\pi \frac{n - n_0}{N}\right) + c \cos\left(4\pi \frac{n - n_0}{N}\right) + d \cos\left(6\pi \frac{n - n_0}{N}\right) \quad (2)$$

with  $a = 0.35875$ ,  $b = 0.48829$ ,  $c = 0.14128$  and  $d = 0.01168$ . The frequency response of the Blackmann-Harris window is shown in Fig. 1. Any other window with a bandlimited frequency response can be applied.

#### 1.3 Inverse FFT synthesis

For mathematical simplicity, the time domain signal is described as a complex signal

$$\tilde{x}_n = w_n \sum_{k=0}^{K-1} A_k \exp\left(2\pi i \omega_k \frac{n - n_0}{N}\right) \quad (3)$$

with

$$A_k = a_k \exp(i\phi_k) \quad (4)$$

of which only the real part of  $\tilde{x}_n$  is retained as the synthesized signal. Evidently, the signal model  $\tilde{x}_n$  has a reciprocal spectrum model  $\tilde{X}_m$ .

$$\tilde{x}_n = \frac{1}{N} \sum_{m=0}^{N-1} \tilde{X}_m \exp(2\pi i m \frac{n - n_0}{N}) \quad (5)$$

where  $\tilde{X}_m$  can be written as

$$\begin{aligned} \tilde{X}_m &= \sum_{n=0}^{N-1} w_n \left[ \sum_{k=0}^{K-1} A_k \exp(2\pi i \omega_k \frac{n - n_0}{N}) \right] \exp(-2\pi i m \frac{n - n_0}{N}) \\ &= \sum_{k=0}^{K-1} A_k \sum_{n=0}^{N-1} w_n \exp(-2\pi i (m - \omega_k) \frac{n - n_0}{N}) \\ &= \sum_{k=0}^{K-1} A_k W(m - \omega_k) \end{aligned} \quad (6)$$

where  $W(m)$  denotes the discrete time fourier transform of  $w_n$ . The spectrum model  $\tilde{X}_m$  is a linear combination of frequency responses of the window, which are shifted over  $\omega_l$  and weighted with a complex factor  $A_l$ . The inverse FFT synthesis algorithm is depicted in Fig. 7.

#### 1.4 Frequency Response Computation

The use of the inverse fast fourier transform (IFFT), requires that the signal length is a power of two. When the window size is chosen to be fixed with a power of two length, an oversampled version of the frequency response can be computed on beforehand and be used as a table look-up (see WO 93/03478).

However, in this invention it is desired to use a variable window length  $M$  which is not necessarily a power of two. The length of the window can be made a power of two again by adding zeros. Three methods are presented to compute the frequency response of a zero padded window being

1. Scaled Table Look-up
2. Analytic Frequency Response Computation
3. Sampled Frequency Response Interpolation

##### 1.4.1 Scaled Table Look-up

The fourier transform of a window with length  $N$  is denoted as  $W^N(m)$ , being

$$W^N(m - n_0) = \sum_{n=0}^{N-1} w^N(n - n_0) \exp\left(-2\pi i \frac{(n - n_0)(m - n_0)}{N}\right) \quad (7)$$

with  $n_0 = \frac{N-1}{2}$ . The window  $w^N(n)$  is now padded up to a length  $P$  which is denoted  $w_N^P(n)$ .

Its frequency response  $W_N^P(m)$  can be expressed in terms of the non padded version  $W^N(m)$  using

$$\begin{aligned} W_N^P(m - p_0) &= \sum_{n=0}^{P-1} w_N^P(n - p_0) \exp\left(-2\pi i \frac{(m - p_0)(n - p_0)}{P}\right) \\ &= \sum_{n=0}^{N-1} w^N(n - n_0) \exp\left(-2\pi i \frac{(m - p_0)N}{P} \frac{(n - n_0)}{N}\right) \\ &= W^N\left(m \frac{N}{P} - n_0\right) \end{aligned} \quad (8)$$

where  $m$  ranges from 1 to  $P$  and  $p_0 = \frac{P-1}{2}$ . This shows that when the frequency response of a window is zero padded, a scaled version of the original frequency response is obtained.

When expressing  $w_N^P(n - p_0)$  as the inverse fourier transform of  $W_N^P(m - p_0)$

$$w_N^P(n - p_0) = \frac{1}{P} \sum_{m=0}^{P-1} W_N^P(m - p_0) \exp\left(2\pi i \frac{(n - p_0)(m - p_0)}{P}\right) \quad (9)$$

and truncating  $W_N^P(m - p_0)$  to a length  $N$ , we obtain  $W_M^N(m - n_0)$ . Its inverse fourier transform  $w_M^N(n - n_0)$  can be written as

$$\begin{aligned} w_M^N(n - n_0) &= \frac{1}{N} \sum_{m=0}^{N-1} W_M^N(m - n_0) \exp\left(2\pi i \frac{(n - n_0)(m - n_0)}{N}\right) \\ &= \frac{1}{N} \sum_{m=0}^{P-1} W_N^P(m - p_0) \exp\left(2\pi i n \frac{(n - n_0)P}{N} \frac{(m - p_0)}{P}\right) \\ &= \frac{P}{N} w_N^P\left(n \frac{P}{N} - p_0\right) \\ &= \frac{N}{M} w^N\left(n \frac{N}{M} - n_0\right) \end{aligned}$$

using

$$\frac{N}{M} = \frac{P}{N}$$

We can now derive

$$\begin{aligned} w_M^N(n - n_0) &= \frac{1}{N} W_M^N(m - n_0) \exp\left(2\pi i \frac{(n - n_0)(m - n_0)}{N}\right) \\ &= \frac{1}{N} W_N^P(m - p_0) \exp\left(2\pi i \frac{(n - n_0)(m - p_0)}{N}\right) \\ &= \frac{1}{N} W^N\left(\frac{N}{P}m - n_0\right) \exp\left(2\pi i \frac{(n - n_0)(m - n_0)}{N}\right) \\ &= \frac{1}{N} W^N\left(\frac{M}{N}m - n_0\right) \exp\left(2\pi i \frac{(n - n_0)(m - n_0)}{N}\right) \end{aligned}$$

what shows explicitly that a window with length  $M$  zero padded up to  $N$  can be computed from the inverse fourier transform of a scaled frequency response  $W^N(m - n_0)$ . This justifies the use of a scaled table look-up of the frequency response. This derivation illustrated by Fig. 3 and the scaled look-up method by Fig. 4.

5 When a window with length  $M$  is zero padded up to a length  $N$ , the main lobe is enlarged up to  $2\frac{N}{M}\beta$ . Therefore, the synthesis of a frequency  $\omega_k$  (see Eq. 6) requires the computation for all frequency domain samples  $m$  for which

$$\begin{aligned} m_{min} &\leq m \leq m_{max} \\ m_{min} &= \lfloor \omega_k - \frac{N}{M}\beta \rfloor \\ m_{max} &= \lceil \omega_k + \frac{N}{M}\beta \rceil \end{aligned} \quad (10)$$

10 From Fig. 1 one can observe that  $\beta = 4$  is a good value for the Blackmann-Harris window. Fig. 2 shows the frequency response of a window which is doubled in length by zero padding. This frequency response is still bandlimited but its bandwidth is doubled. When comparing this with figure 1, it is clear that this is the same frequency response, scaled with a factor 2.

15 Conclusion: As a first method for the frequency response computation, the invention uses a look-up table containing an oversampled frequency response  $W^N(m)$  of the main lobe and a scaling factor  $\frac{M}{N}$  yielding  $W^N(m\frac{M}{N})$ . The method is illustrated in Fig. 4. This method is very fast but its accuracy is dependent on the size of the look-up table and might not be the best solution when strict memory constraints must be taken into account or when very  
20 high accuracy is required.

#### 1.4.2 Analytic Frequency Response Computation

In the case that memory constraints prohibit the use of a look-up table or when a very high accuracy is desired it is appropriate to compute the analytic form of the frequency response function. For every sample  $m$ , the response is still computed in constant time but  
25 this is clearly slower then the scaled table look-up method. The frequency response of a zero

padded Blackmann-Harris window is now computed.

$$\begin{aligned}
 & W_M^N(m) \\
 = & \sum_{n=0}^{M-1} w_n \exp(-2\pi i m \frac{n - n_0}{N}) \\
 = & \sum_{n=0}^{M-1} \left[ a + \frac{b}{2} \left( \exp(2\pi i \frac{n - n_0}{M}) + \exp(-2\pi i \frac{n - n_0}{M}) \right) + \right. \\
 & \quad \frac{c}{2} \left( \exp(4\pi i \frac{n - n_0}{M}) + \exp(-4\pi i \frac{n - n_0}{M}) \right) + \\
 & \quad \left. \frac{d}{2} \left( \exp(6\pi i \frac{n - n_0}{M}) + \exp(-6\pi i \frac{n - n_0}{M}) \right) \right] \exp(-2\pi i m \frac{n - n_0}{N}) \\
 = & \sum_{n=0}^{M-1} \left[ a \exp(-2\pi i(n - n_0) \frac{m}{N}) + \right. \\
 & \quad \frac{b}{2} \left( \exp(2\pi i \frac{n - n_0}{M}(1 - \frac{Mm}{N})) + \exp(-2\pi i \frac{n - n_0}{M}(1 + \frac{Mm}{N})) \right) + \\
 & \quad \frac{c}{2} \left( \exp(2\pi i \frac{n - n_0}{M}(2 - \frac{Mm}{N})) + \exp(-2\pi i \frac{n - n_0}{M}(2 + \frac{Mm}{N})) \right) + \\
 & \quad \left. \frac{d}{2} \left( \exp(2\pi i \frac{n - n_0}{M}(3 - \frac{Mm}{N})) + \exp(-2\pi i \frac{n - n_0}{M}(3 + \frac{Mm}{N})) \right) \right] \\
 \end{aligned} \tag{11}$$

10

Since  $n_0 = \frac{N-1}{2}$  and using the equality

$$\sum_{n=0}^{N-1} x^{n-\frac{N-1}{2}} = x^{-\frac{N-1}{2}} \sum_{n=0}^{N-1} x^n = \frac{x^{-\frac{N}{2}}}{x^{-\frac{1}{2}}} \frac{x^N - 1}{x - 1} = \frac{x^{\frac{N}{2}} - x^{-\frac{N}{2}}}{x^{\frac{1}{2}} - x^{-\frac{1}{2}}}$$

with  $x = \exp(2\pi i m/N)$ , it is obtained that

15

$$\sum_{n=0}^{N-1} \exp(2\pi i m \frac{n - n_0}{N}) = \frac{\sin(\pi \frac{mM}{N})}{\sin(\pi \frac{m}{N})}$$

resulting in

20

$$\begin{aligned}
 & W_M^N(m) \\
 = & a \frac{\sin(\pi \frac{mM}{N})}{\sin(\pi \frac{m}{N})} + \\
 & \frac{b}{2} \left[ \frac{\sin(\pi(\frac{mM}{N} - 1))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} - 1))} + \frac{\sin(\pi(\frac{mM}{N} + 1))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} + 1))} \right] + \\
 & \frac{c}{2} \left[ \frac{\sin(\pi(\frac{mM}{N} - 2))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} - 2))} + \frac{\sin(\pi(\frac{mM}{N} + 2))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} + 2))} \right] + \\
 & \frac{d}{2} \left[ \frac{\sin(\pi(\frac{mM}{N} - 3))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} - 3))} + \frac{\sin(\pi(\frac{mM}{N} + 3))}{\sin(\pi \frac{1}{M}(\frac{mM}{N} + 3))} \right]
 \end{aligned} \tag{13}$$

When no zero padding is applied, meaning that  $M = N$ , this can be simplified further to

$$\begin{aligned}
 & W^N(m) \\
 &= a \frac{\sin(\pi m)}{\sin(\pi \frac{m}{N})} + \\
 &\quad \frac{b}{2} \left[ \frac{\sin(\pi(m-1))}{\sin(\pi \frac{1}{N}(m-1))} + \frac{\sin(\pi(m+1))}{\sin(\pi \frac{1}{N}(m+1))} \right] + \\
 &\quad \frac{c}{2} \left[ \frac{\sin(\pi(m-2))}{\sin(\pi \frac{1}{N}(m-2))} + \frac{\sin(\pi(m+2))}{\sin(\pi \frac{1}{N}(m+2))} \right] + \\
 &\quad \frac{d}{2} \left[ \frac{\sin(\pi(m-3))}{\sin(\pi \frac{1}{N}(m-3))} + \frac{\sin(\pi(m+3))}{\sin(\pi \frac{1}{N}(m+3))} \right]
 \end{aligned} \tag{14}$$

5

Care must be taken that the denominator is not too close to zero. When this is the case, every exponential in the sum series yields 1, resulting in  $N$  for the sum. Hence, in a small interval  $-\epsilon < m < \epsilon$ , the expression

10

$$\frac{\sin(2\pi m)}{\sin(2\pi m/N)} \tag{15}$$

returns  $N$ .

In addition, the number of sine computations can be reduced by using the following trigonometric equalities

15

$$\begin{aligned}
 \sin(\alpha + \beta) &= \cos(\alpha) \sin(\beta) + \sin(\alpha) \cos(\beta) \\
 \cos(\alpha + \beta) &= \cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta)
 \end{aligned}$$

For instance, expressing

$$\begin{aligned}
 \sin(\pi(m-2)) &= \sin(\pi(m-3) + \pi) \\
 \sin(\pi \frac{1}{M}(m-2)) &= \sin(\pi(m-3) + \frac{\pi}{M})
 \end{aligned}$$

shows that all terms in equation (14) can be computed recursively from  $\sin(\pi(\frac{mM}{N}-3))$  and  
20  $\sin(\frac{\pi}{M}(\frac{mM}{N}-3))$ .

Again, since

$$\begin{aligned}
 \sin(\pi(\frac{(m+1)M}{N}-3)) &= \sin(\pi(\frac{mM}{N}-3) + \frac{\pi M}{N}) \\
 \sin(\pi \frac{1}{M}(\frac{(m+1)M}{N}-3)) &= \sin(\pi(\frac{mM}{N}-3) + \frac{\pi}{N})
 \end{aligned}$$

25 the recursion is applied for computing the next sample  $m+1$  from values that were used for computing the sample  $m$ .

Table 1: Coefficients for different windows.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
Blackmann-Harris	0.35875	0.48829	0.14128	0.01168
Blackmann	0.42	0.5	0.08	0
Hamming	0.54	0.46	0	0
Hanning	0.5	0.5	0	0

Note that this solution is applicable to all windows that have a similar form as Eq. (2). They are listed in table 1.

The computation of the analytic frequency response is illustrated in Fig. 5.

#### 1.4.3 Sampled Frequency Response Interpolation

Finally, when the table look-ups cannot be used and the analytic frequency response of the window cannot be computed, a third method can be used which is named Sampled Frequency Interpolation. Again, it computes the response for a frequency response in constant time but is less efficient than the previous methods. The non integer shift of  $W(m)$  over  $\omega_l$  is computed using a sampled spectrum  $W_k$  obtained by an FFT of the zero padded window. Note that also here, the division by zero must be avoided. This method is illustrated in Fig. 6.

$$\begin{aligned}
 & W(m) \\
 &= \sum_{n=0}^{N-1} w_n \exp(-2\pi i m \frac{n - n_0}{N}) \\
 &= \sum_{n=0}^{N-1} \left[ \frac{1}{N} \sum_{k=0}^{N-1} W_k \exp(2\pi i k \frac{n - n_0}{N}) \right] \exp(-2\pi i m \frac{n - n_0}{N}) \\
 &= \sum_{k=0}^{N-1} W_k \sum_{n=0}^{N-1} \exp(2\pi i (k - m) \frac{n - n_0}{N}) \\
 &= \sum_{k=0}^{N-1} W_k \frac{\sin(2\pi(k - m))}{\sin(2\pi(k - m)/N)}
 \end{aligned} \tag{16}$$

## 2 Amplitude Computation

### 2.1 Introduction

In this section, an efficient least mean squares technique is described for the computation of the amplitudes and phases. In WO 90/13887, the estimation of the amplitudes is claimed by detecting individual peaks in the magnitude spectrum, and performing a parabolic interpolation to refine the frequency and amplitude values. In WO 93/04467 and WO 95/30983 a least means squares is presented which is applied iteratively on the signal, subtracting a single sinusoidal component each time.

The major difference with the present invention is that all amplitudes and phases are computed simultaneously for a given set of frequencies. This allows to resolve strongly overlapping frequency responses of sinusoidal components. As will be shown later, the original computational complexity of this method is  $\mathcal{O}(K^2N)$  where the  $K$  denotes the number of partials and  $N$  the number of samples. The invention however, solves this problem in  $\mathcal{O}(N \log(N))$  and reduces the space complexity, which is originally  $\mathcal{O}(K^2)$ , to  $\mathcal{O}(K)$ . The amplitude computation method is illustrated in Figure 8.

## 2.2 Complex Amplitude Computation

We describe how the amplitudes and phases are computed for all sinusoidal components. We reformulate Eq. (1) as a sum of cosines and sines where the real part of the complex amplitude is denoted  $A_l^r = a_l \cos \phi_l$  and the imaginary part as  $A_l^i = a_l \sin \phi_l$ . The signal model for the short time signal  $\tilde{x}_n$  can now be written as

$$\begin{aligned} \tilde{x}_n &= w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n - n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n - n_0}{N}) \right) \\ &= w_n \sum_{k=0}^{K-1} \left( A_k^r \cos(2\pi \omega_k \frac{n - n_0}{N}) - A_k^i \sin(2\pi \omega_k \frac{n - n_0}{N}) \right) \end{aligned} \quad (17)$$

The error function  $\chi(\bar{A}; \bar{\omega})$  expresses the square difference between the samples in the windowed signal  $x_n$  and the signal model  $\tilde{x}_n$ .

$$\chi(\bar{A}; \bar{\omega}) = \sum_n (x_n - \tilde{x}_n)^2 \quad (18)$$

The notation indicates that the error is minimized with respect to a vector of variables  $\bar{A}$  for a given set of frequencies  $\bar{\omega}$  that are assumed to be known. The minimization is realized by putting the derivatives with respect to the unknowns to zero

$$\frac{\partial \chi(\bar{A}; \bar{\omega})}{\partial A_l^r} = 0, \frac{\partial \chi(\bar{A}; \bar{\omega})}{\partial A_l^i} = 0 \quad (19)$$

resulting respectively in

$$\begin{aligned}
 & \sum_{k=0}^{K-1} A_k^r \left( \sum_{n=0}^{N-1} w_n^2 \cos\left(2\pi\omega_k \frac{n-n_0}{N}\right) \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right) \right) \\
 & - \sum_{k=0}^{K-1} A_k^i \left( \sum_{n=0}^{N-1} w_n^2 \sin\left(2\pi i\omega_k \frac{n-n_0}{N}\right) \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right) \right) \\
 & = \sum_{n=0}^{N-1} x_n w_n \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right)
 \end{aligned} \tag{20}$$

and

$$\begin{aligned}
 & - \sum_{k=0}^{K-1} A_k^r \left( \sum_{n=0}^{N-1} w_n^2 \cos\left(2\pi\omega_k \frac{n-n_0}{N}\right) \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right) \right) \\
 & + \sum_{k=0}^{K-1} A_k^i \left( \sum_{n=0}^{N-1} w_n^2 \sin\left(2\pi i\omega_k \frac{n-n_0}{N}\right) \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right) \right) \\
 & = - \sum_{n=0}^{N-1} x_n w_n \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right)
 \end{aligned} \tag{21}$$

These two sets of  $K$  equations have  $2K$  unknown variables what can be written in the following matrix form

$$\begin{bmatrix} \mathbf{B}^{1,1} & \mathbf{B}^{1,2} \\ \mathbf{B}^{2,1} & \mathbf{B}^{2,2} \end{bmatrix} \begin{bmatrix} \mathbf{A}^r \\ \mathbf{A}^i \end{bmatrix} = \begin{bmatrix} \mathbf{C}^1 \\ \mathbf{C}^2 \end{bmatrix} \tag{22}$$

with

$$\begin{aligned}
 B_{l,k}^{1,1} &= \sum_{n=0}^{N-1} w_n^2 \cos\left(2\pi\omega_k \frac{n-n_0}{N}\right) \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right) \\
 B_{l,k}^{1,2} &= - \sum_{n=0}^{N-1} w_n^2 \sin\left(2\pi\omega_k \frac{n-n_0}{N}\right) \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right) \\
 B_{l,k}^{2,1} &= - \sum_{n=0}^{N-1} w_n^2 \cos\left(2\pi\omega_k \frac{n-n_0}{N}\right) \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right) \\
 B_{l,k}^{2,2} &= \sum_{n=0}^{N-1} w_n^2 \sin\left(2\pi\omega_k \frac{n-n_0}{N}\right) \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right) \\
 C_l^1 &= \sum_{n=0}^{N-1} x_n w_n \cos\left(2\pi\omega_l \frac{n-n_0}{N}\right) \\
 C_l^2 &= - \sum_{n=0}^{N-1} x_n w_n \sin\left(2\pi\omega_l \frac{n-n_0}{N}\right)
 \end{aligned}$$

15

Under the condition that every sinusoid has a different frequency, the matrix  $\mathbf{B}$  cannot have two linear dependent rows implying a unique solution for  $\mathbf{A}$ .

The computational complexity of this method is very high, for instance,

- the computation of the matrix  $\mathbf{B}$  has a complexity  $\mathcal{O}(K^2N)$
- the computation of the matrix  $\mathbf{C}$  has a complexity  $\mathcal{O}(KN)$
- the solution of the linear set of equations is  $\mathcal{O}(K^3)$

5 Note that the order of magnitude of  $K$  and  $N$  is not significantly different. In the next sections, the complexity is reduced to  $\mathcal{O}(N \log(N))$ .

### 2.3 Efficient Computation of $\mathbf{B}$

Several optimizations for previous computation are disclosed. The main computational burden comes from the construction of the matrix  $\mathbf{B}$  and the solution of the system of linear 10 equations which have complexity  $\mathcal{O}(K^2N)$  and  $\mathcal{O}(K^3)$  respectively. Following derivation shows that this can be improved considerably. We start by writing

$$\begin{aligned} B_{l,k}^{1,1} &= \sum_{n=0}^{N-1} w_n^2 \cos\left(2\pi\omega_k \frac{n - n_0}{N}\right) \cos\left(2\pi\omega_l \frac{n - n_0}{N}\right) \\ &= \frac{1}{2} \sum_{n=0}^{N-1} w_n^2 \left[ \cos\left(2\pi(\omega_k + \omega_l) \frac{n - n_0}{N}\right) + \cos\left(2\pi(\omega_k - \omega_l) \omega \frac{n - n_0}{N}\right) \right] \\ &= \frac{1}{2} (\Re(Y(\omega_k + \omega_l)) + \Re(Y(\omega_k - \omega_l))) \end{aligned} \quad (23)$$

15 with

$$Y(m) = \frac{1}{2} \sum_{n=0}^{N-1} w_n^2 \left( \exp\left(2\pi i m \frac{n - n_0}{N}\right) \right) \quad (24)$$

which is the time inverted fourier transform of the square window. In an analogue manner one obtains

$$\begin{aligned} B_{l,k}^{1,2} &= -\frac{1}{2} (\Im(Y(\omega_k + \omega_l)) + \Im(Y(\omega_k - \omega_l))) \\ B_{l,k}^{2,1} &= -\frac{1}{2} (\Im(Y(\omega_k + \omega_l)) - \Im(Y(\omega_k - \omega_l))) \\ B_{l,k}^{2,2} &= -\frac{1}{2} (\Re(Y(\omega_k + \omega_l)) - \Re(Y(\omega_k - \omega_l))) \end{aligned} \quad (25)$$

In figure 2 the frequency response of the squared Blackmann-Harris window is shown, which is clearly still bandlimited. This can be understood easily, considering that taking the square in the time domain is equivalent with a convolution in the frequency domain.

Since the window is real and symmetric, its frequency response is also real and symmetric. This means that  $B^{1,2}$  and  $B^{2,1}$  only contain zeros. When defining the matrix  $Y_{-,l,k}$  as  $Y(\omega_k - \omega_l)$  and the matrix  $Y_{+,l,k}$  as  $Y(\omega_k + \omega_l)$  one obtains

$$B^{1,1} = \frac{1}{2}(Y^+ + Y^-) \quad (26)$$

$$B^{2,2} = -\frac{1}{2}(Y^+ - Y^-) \quad (27)$$

$$Y^- = \begin{bmatrix} Y(0) & Y(\omega_1 - \omega_0) & Y(\omega_2 - \omega_0) & Y(\omega_3 - \omega_0) \\ Y(\omega_0 - \omega_1) & Y(0) & Y(\omega_2 - \omega_1) & Y(\omega_2 - \omega_1) \\ Y(\omega_0 - \omega_2) & Y(\omega_1 - \omega_2) & Y(0) & Y(\omega_3 - \omega_2) \\ Y(\omega_0 - \omega_3) & Y(\omega_1 - \omega_3) & Y(\omega_2 - \omega_3) & Y(0) \end{bmatrix}$$

$$Y^+ = \begin{bmatrix} Y(2\omega_0) & Y(\omega_1 + \omega_0) & Y(\omega_2 + \omega_0) & Y(\omega_3 + \omega_0) \\ Y(\omega_0 + \omega_1) & Y(2\omega_1) & Y(\omega_1 + \omega_2) & Y(\omega_1 + \omega_3) \\ Y(\omega_0 + \omega_2) & Y(\omega_1 + \omega_2) & Y(2\omega_2) & Y(\omega_3 + \omega_2) \\ Y(\omega_0 + \omega_3) & Y(\omega_1 + \omega_3) & Y(\omega_2 + \omega_3) & Y(2\omega_3) \end{bmatrix}$$

for a single harmonic sound source, this can be simplified further to

$$Y^- = \begin{bmatrix} Y(0) & Y(\omega) & Y(2\omega) & Y(3\omega) \\ Y(-\omega) & Y(0) & Y(\omega) & Y(2\omega) \\ Y(-2\omega) & Y(-\omega) & Y(0) & Y(\omega) \\ Y(-3\omega) & Y(-2\omega) & Y(-\omega) & Y(0) \end{bmatrix}$$

$$Y^+ = \begin{bmatrix} Y(0) & Y(\omega) & Y(2\omega) & Y(3\omega) \\ Y(\omega) & Y(2\omega) & Y(3\omega) & Y(4\omega) \\ Y(2\omega) & Y(3\omega) & Y(4\omega) & Y(5\omega) \\ Y(3\omega) & Y(4\omega) & Y(5\omega) & Y(6\omega) \end{bmatrix}$$

where  $\omega$  is the fundamental frequency.

The crucial observation that has to be made in this invention is that the matrix  $B$  becomes a band diagonal matrix when the sinusoidal components are sorted by their frequency. If the components are sorted, the frequency differences for elements close to the diagonal of  $Y^-$  is small and will fall in the main lobe of  $Y$  yielding a large value. For the elements far from the diagonal, the frequency difference is large and will fall outside the main lobe of  $Y$ . For  $Y^+$  a similar reasoning can be applied. Elements close to the upper left corner have the smallest

frequency and can fall into the main lobe. The elements in the lower right corner have a very large value and can fall into the main lobe because of spectral replication. As a result, matrices  $\mathbf{B}^{1,1}$  and  $\mathbf{B}^{2,2}$  are band diagonal which has a major impact on the computation time that is required to solve the equations.

5 A typical method to solve a linear set of equations is the use of Gaussian elimination with backsubstitution. This method has a time complexity  $\mathcal{O}(K^3)$ . However, since the system is band diagonal, this method requires only a linear time complexity  $\mathcal{O}(K)$ .

In addition, the space complexity can be reduced from  $\mathcal{O}(K^2)$  to  $\mathcal{O}(K)$  by storing only the diagonal band. Let us define a shifted matrix

$$10 \quad \overleftarrow{\mathbf{B}}_{l,k} = B_{l,l+k-D} \quad (28)$$

where  $D$  denotes the number of diagonals that are stored around the main diagonal. Note that  $l = 0, \dots, L-1$  and  $k = 0, \dots, 2D$ . For combinations  $(k, l)$  resulting in an index outside  $\mathbf{A}$ , a zero value is returned. The amplitudes are computed directly from the shifted versions of  $\mathbf{B}^{1,1}$ ,  $\mathbf{B}^{2,2}$ . By denoting this routine as *SOLVE* this is written as

$$15 \quad \begin{aligned} \mathbf{A}^r &= \text{SOLVE}(\overleftarrow{\mathbf{B}}^{1,1}, \mathbf{C}^1) \\ \mathbf{A}^i &= \text{SOLVE}(\overleftarrow{\mathbf{B}}^{2,2}, \mathbf{C}^2) \end{aligned} \quad (29)$$

### Conclusions:

- The original complexity for the computation of  $\mathbf{B}$  was  $\mathcal{O}(K^2N)$
- Using one of the constant time frequency response computation methods that were discussed in the synthesis section, this can be reduced to  $\mathcal{O}(K^2)$
- Since  $\mathbf{B}$  is band diagonal, only elements that are close to the diagonal must be computed resulting in  $\mathcal{O}(K)$
- A second result of the band diagonal form of  $\mathbf{B}$  is that the system can now be solved in  $\mathcal{O}(K)$  instead of  $\mathcal{O}(K^3)$
- Finally, also the space complexity of  $\mathbf{B}$  is reduced from  $\mathcal{O}(K^2)$  to  $\mathcal{O}(K)$

This computation is illustrated in Figure 8.

#### 2.4 Efficient Computation of $\mathbf{C}$

The results of the previous section moves the bottleneck to the computation of  $\mathbf{C}$  which has a complexity  $\mathcal{O}(KN)$ .

$$\begin{aligned}
 C_l &= \sum_{n=0}^{N-1} x_n w_n \exp\left(2\pi i \omega_l \frac{n - n_0}{N}\right) \\
 &= \sum_{n=0}^{N-1} \left[ \sum_{m=0}^{N-1} X_m \exp\left(2\pi i m \frac{n}{N}\right) \right] w_n \exp\left(2\pi i \omega_l \frac{n}{N}\right) \\
 &= \sum_{m=0}^{N-1} X_m W(m + \omega_l)
 \end{aligned} \tag{30}$$

5

By taking the real and imaginary part respectively,  $\mathbf{C}^1$  and  $\mathbf{C}^2$  are obtained. Analogue to Eq. (10), only samples of the main lobe must be computed, yielding

$$C_l = \sum_{m=m_{\min}}^{m_{\max}} X_m W(m + \omega_l) \tag{31}$$

Again, the frequency response  $W(m)$  can be computed by previously described methods.

10 The fourier transform of  $x_n$  however requires a time complexity  $\mathcal{O}(N \log N)$ .

This computation is illustrated in Figure 8.

## 2.5 Amplitude Computation Pre-processing

The goal of the pre-processing before the amplitude computation is twofold. On one hand the frequencies are sorted in order to obtain a band diagonal matrix for  $\mathbf{B}$ . In addition, 15 frequencies that occur twice result in two exact rows in  $\mathbf{B}$  making it a singular matrix. Therefore, no double frequencies are allowed for the frequency computation.

On the other hand, the preprocessing determines how many diagonals of the matrix  $\mathbf{B}$  must be taken into account. This is done by counting the number of sinusoidal components that fall in the main lobe of each frequency response. The maximum number of components 20 over all frequency responses yields the value for  $D$ .

The amplitude estimation preprocessing is depicted in Figure 9.

## 3 Frequency Optimization

### 3.1 Introduction

The previous sections described an efficient method to compute the amplitudes and phases 25 for a given set of frequencies. In this part of the invention, methods are presented to get the initial values for the frequencies and methods to optimize these frequencies with respect

to the error criterium in an iterative manner. This optimization requires the computation of the gradient and hessian of the error function which normally require time complexity  $\mathcal{O}(K^2N)$ . The invention discloses a method that reduces the computational complexity to  $\mathcal{O}(N \log(N))$ . Note that still a variable window length is used.

5 For the optimization of the frequencies two types of signal models are distinguished. One containing arbitrary frequency components, as given previously by

$$\tilde{x}_n = w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n - n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n - n_0}{N}) \right) + r_n$$

On the other hand when the sound sources are known to be pitched, a model of harmonic spectra can be used which is written as

$$10 \quad \tilde{x}_n = w_n \frac{1}{2} \sum_{k=0}^{K-1} \sum_{q=1}^{Q_k-1} \left( A_{k,q} \exp(2\pi i q \omega_k \frac{n - n_0}{N}) + A_{k,q}^* \exp(-2\pi i q \omega_k \frac{n - n_0}{N}) \right) + r_n \quad (32)$$

In this last case, the model consists of  $K$  sources each modelled by  $Q_k$  harmonic components. For the first model, all frequencies must be optimized while for the second model, only the fundamental frequencies are optimized. As a result, a different frequency optimization is disclosed for each model. The amplitude estimation on the other hand, stays the same for 15 both models.

### 3.2 Initial Frequency Values

The error function is very non linear in function of the frequencies and can therefore get stuck in local minima. Therefore, the initial values of the frequencies plays an important role. Different methods can be used

20 • **(Multi)pitch estimation:** For the harmonic signal model (Eq. 32), any (multi)pitch estimator can be used to compute a number of pitches from the original signal. From these pitches, sets of frequencies can be produced depending on prior knowledge of the sound source. For instance a harmonic series of frequencies for periodic sounds, a inharmonic frequency series for strings etc. Also by matching the spectrum to a database of spectra that are labelled with a series of frequencies the initial values can 25 be obtained.

• **Peak picking.** For the non harmonic model, individual local maxima can be detected in the sampled spectrum obtained by an FFT. Note that in that case the analysis window must be sufficiently large.

- Prior knowledge. The frequencies and / or pitches can also be read from an external file such as a musical score or an annotation of the signal.

This is depicted in Figures 10.

### 3.3 Local Quadratic Approximation

5 The problem of minimizing continuous differential functions of many variables is widely studied and any of the conventional methods can be applied. Many of these methods make a local quadratic approximation of the error function. Let us consider a second order Taylor expansion around the error function  $\chi$ , only this time, the amplitudes are kept fixed and the variation in function of the frequencies  $\omega$  is studied. This is denoted  $\chi(\bar{\omega}; \bar{A})$ . The expansion  
10 of  $\chi(\bar{\omega}; \bar{A})$  around a given value  $\bar{v}$  results in a local quadratic approximation denoted as

$$\begin{aligned}\chi(\bar{\omega}; \bar{A}) &\simeq \chi(\bar{v}; \bar{A}) + (\bar{\omega} - \bar{v})^T \mathbf{h}|_{\bar{v}} + \frac{1}{2}(\bar{\omega} - \bar{v})^T \mathbf{H}|_{\bar{v}}(\bar{\omega} - \bar{v}) \\ \nabla \chi(\bar{\omega}; \bar{A}) &\simeq \mathbf{h}|_{\bar{v}} + \mathbf{H}|_{\bar{v}}(\bar{\omega} - \bar{v})\end{aligned}\quad (33)$$

where  $\mathbf{h}|_{\bar{v}}$  is the gradient and  $\mathbf{H}|_{\bar{v}}$  the Hessian matrix of  $\chi$  evaluated at  $\bar{v}$ .

$$\begin{aligned}h_k|_{\bar{v}} &\equiv \frac{\partial \chi(\bar{\omega}; \bar{A})}{\partial \bar{\omega}_k}|_{\bar{v}} \\ H_{l,k}|_{\bar{v}} &\equiv \frac{\partial^2 \chi(\bar{\omega}; \bar{A})}{\partial \bar{\omega}_l \partial \bar{\omega}_k}|_{\bar{v}}\end{aligned}$$

15

When  $\bar{v}$  is a minimum of the error function, the gradient at  $\bar{v}$  is zero so that Eq. (33) becomes

$$\begin{aligned}\chi(\bar{\omega}; \bar{A}) &\simeq \chi(\bar{v}; \bar{A}) + \frac{1}{2}(\bar{\omega} - \bar{v})^T \mathbf{H}|_{\bar{v}}(\bar{\omega} - \bar{v}) \\ \nabla \chi(\bar{\omega}; \bar{A}) &\simeq \mathbf{H}|_{\bar{v}}(\bar{\omega} - \bar{v})\end{aligned}\quad (34)$$

Note that for the local approximation, the gradient is linear in function of  $\bar{\omega}$  and the Hessian  
20 independent of  $\bar{\omega}$  from which follows that  $\mathbf{H}|_{\bar{v}}$  is equivalent with  $\mathbf{H}|_{\bar{\omega}}$ . For a given  $\bar{\omega}$  the minimum  $\bar{v}$  can now be approximated using

$$\bar{v} \simeq \bar{\omega} - \mathbf{H}|_{\bar{\omega}}^{-1} \nabla \chi(\bar{\omega}; \bar{A}) \quad (35)$$

Two major classes of optimization algorithms can be distinguished. The first class are the  
25 gradient descent methods which use first order information of the error function in order to optimize the frequency values

$$\bar{\omega}^{(r)} = \bar{\omega}^{(r-1)} - \eta \mathbf{h}|_{\bar{\omega}^{r-1}} \quad (36)$$

The superscript  $(r)$  denotes the index of the iteration and  $\eta$  the learning rate. Many different variations have been proposed such as using a momentum term, line search algorithms, using conjugate gradients and scaled conjugate gradients.

From Eq. 34 follows that the error function can be optimized using

$$^5 \quad \bar{\omega}^{(r)} = \bar{\omega}^{(r-1)} - H_{|\bar{\omega}^{(r-1)}}^{-1} h_{|\bar{\omega}^{(r-1)}} \quad (37)$$

This is called Newton's method. Because of the large computational cost for the inversion of the Hessian of the matrix, quasi-Newton methods have been developed which construct the inverse matrix iteratively. In the present invention however, it is shown that the hessian is a band diagonal matrix of which the inverse can be computed in linear time.

<sup>10</sup> Finally, following termination criteria can be used to stop the iterative optimization of the frequencies.

- stop after a fixed number of iterations
- stop after fixed computation time
- stop when error function drops below a specified value
- <sup>15</sup> • stop when the error change drops below a specified value
- stop when error measure starts to increase.

### 3.4 Efficient Computation of the Gradient

An efficient method is disclosed which allows to compute the gradient of the error function with respect to the frequencies in  $\mathcal{O}(N \log(N))$  time. Each element of the gradient is <sup>20</sup> computed in constant time, implying  $\mathcal{O}(K)$  for all elements. However, the FFT of the noise residual  $R_m$  is required implying  $\mathcal{O}(N \log(N))$ . The gradient of the error function for the

non harmonic model results in

$$\begin{aligned}
 & \frac{\partial \chi(\bar{\omega}; \bar{A})}{\partial \omega_l} \\
 &= \frac{\partial}{\partial \omega_l} \sum_{n=0}^{N-1} \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right]^2 \\
 &= \sum_{n=0}^{N-1} 2 \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right] \\
 &\quad \left( -w_n \frac{1}{2} A_l \exp(2\pi i \omega_l \frac{n-n_0}{N}) 2\pi i \frac{n-n_0}{N} - w_n \frac{1}{2} A_l^* \exp(-2\pi i \omega_l \frac{n-n_0}{N}) \left( -2\pi i \frac{n-n_0}{N} \right) \right) \\
 &= \sum_{n=0}^{N-1} \left[ \sum_{m=0}^{N-1} \frac{1}{N} R_m \exp(2\pi i m \frac{n-n_0}{N}) \right] \\
 &\quad \left( -w_n A_l \exp(2\pi i \omega_l \frac{n-n_0}{N}) 2\pi i \frac{n-n_0}{N} + w_n A_l^* \exp(-2\pi i \omega_l \frac{n-n_0}{N}) 2\pi i \frac{n-n_0}{N} \right) \\
 &= \frac{1}{N} \sum_{m=0}^{N-1} R_m \left[ -A_l \sum_{n=0}^{N-1} w_n 2\pi i \frac{n-n_0}{N} \exp(2\pi i(m+\omega_l) \frac{n-n_0}{N}) \right. \\
 &\quad \left. + A_l^* \sum_{n=0}^{N-1} w_n 2\pi i \frac{n-n_0}{N} \exp(2\pi i(m-\omega_l) \frac{n-n_0}{N}) \right] \\
 &= \frac{1}{N} \sum_{m=0}^{N-1} R_m (A_l^* W'(m-\omega_l) - A_l W'(m+\omega_l)) \\
 &= \frac{2}{N} \sum_{m=0}^{N-1} \Re(R_m A_l^* W'(m-\omega_l)) \\
 &= \frac{2}{N} \sum_{m=m_{min}}^{m_{max}} \Re(R_m A_l^* W'(m-\omega_l)) \tag{38}
 \end{aligned}$$

where  $m_{min}$  and  $m_{max}$  are dependent of  $\omega_l$  and are computed from the bandwidth  $\beta$  of  $W'(m)$  as in Eq. (10).

For the harmonic model we obtain

$$\begin{aligned}
 & \frac{\partial \chi(\bar{\omega}; \bar{A})}{\partial \omega_l} \\
 = & \frac{\partial}{\partial \omega_l} \sum_{n=0}^{N-1} \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \sum_{q=1}^{Q_k-1} \left( A_{k,q} \exp(2\pi i q \omega_k \frac{n-n_0}{N}) + A_{k,q}^* \exp(-2\pi i q \omega_k \frac{n-n_0}{N}) \right) \right]^2 \\
 = & \sum_{n=0}^{N-1} 2 \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \sum_{q=1}^{Q_k-1} \left( A_{k,q} \exp(2\pi i q \omega_k \frac{n-n_0}{N}) + A_{k,q}^* \exp(-2\pi i q \omega_k \frac{n-n_0}{N}) \right) \right] \\
 & \left( -w_n \frac{1}{2} \sum_{q=1}^{Q_l-1} q A_{l,q} 2\pi i \frac{n-n_0}{N} \exp(2\pi i q \omega_l \frac{n-n_0}{N}) \right. \\
 & \quad \left. - w_n \frac{1}{2} \sum_{q=1}^{Q_l-1} q A_{l,q}^* 2\pi i \frac{n-n_0}{N} \exp(-2\pi i q \omega_l \frac{n-n_0}{N}) \right) \\
 = & \sum_{n=0}^{N-1} \left[ \sum_{m=0}^{N-1} \frac{1}{N} R_m \exp(2\pi i m \frac{n-n_0}{N}) \right] \\
 & \left( -w_n \sum_{q=1}^{Q_l-1} q A_{l,q} 2\pi i \frac{n-n_0}{N} \exp(2\pi i q \omega_l \frac{n-n_0}{N}) \right. \\
 & \quad \left. - w_n \sum_{q=1}^{Q_l-1} q A_{l,q}^* 2\pi i \frac{n-n_0}{N} \exp(-2\pi i q \omega_l \frac{n-n_0}{N}) \right) \\
 = & \frac{1}{N} \sum_{m=0}^{N-1} R_m \left[ - \sum_{q=1}^{Q_l-1} q A_{l,q} \sum_{n=0}^{N-1} w_n 2\pi i \frac{n-n_0}{N} \exp(2\pi i(m+q\omega_l) \frac{n-n_0}{N}) \right. \\
 & \quad \left. + \sum_{q=1}^{Q_l-1} q A_{l,q}^* \sum_{n=0}^{N-1} w_n 2\pi i \frac{n-n_0}{N} \exp(2\pi i(m-q\omega_l) \frac{n-n_0}{N}) \right] \\
 = & \frac{1}{N} \sum_{m=0}^{N-1} R_m \sum_{q=1}^{Q_l-1} (q A_{l,q}^* W'(m-q\omega_l) - q A_{l,q} W'(m+q\omega_l)) \\
 = & \frac{2}{N} \sum_{q=1}^{Q_l-1} \sum_{m=m_{min}}^{m_{max}} \Re(R_m q A_{l,q}^* W'(m-q\omega_l)) \tag{39}
 \end{aligned}$$

Note that,

$$\begin{aligned}
 R_m &= \sum_{n=0}^{N-1} r_n \exp(2\pi i m \frac{n}{N}) \\
 W(m) &= \sum_{n=0}^{N-1} w_n \exp(2\pi i m \frac{n-n_0}{N}) \\
 W'(m) &= \frac{\partial W(m)}{\partial m} = \sum_{n=0}^{N-1} w_n 2\pi i \frac{n-n_0}{N} \exp(2\pi i m \frac{n-n_0}{N})
 \end{aligned}$$

Hence,  $W'(m)$  is the first fourier derivative of  $W(m)$  which still results in a bandlimited frequency response as is shown in Fig.1.

### 3.5 Efficient Computation of the Hessain Matrix

In this section it is derived that the Hessian matrix for the non harmonic signal model is band diagonal. An efficient method is disclosed which computes each element of the Hessian in constant time. The computation of all band diagonal elements is computed in  $\mathcal{O}(K)$  time. Since the fourier transform of the noise residual  $R_m$  is required the total complexity is  $\mathcal{O}(N \log(N))$

The Hessian for the non harmonic model is computed

$$\begin{aligned}
 & \frac{\partial \chi(\bar{A}; \bar{\omega})}{\partial \omega_p \partial \omega_l} \\
 &= \frac{\partial}{\partial \omega_p} \frac{\partial}{\partial \omega_l} \sum_{n=0}^{N-1} \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right]^2 \\
 &= \frac{\partial}{\partial \omega_p} \sum_{n=0}^{N-1} 2 \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right] \\
 &\quad \left( -w_n \frac{1}{2} A_l \exp(2\pi i \omega_l \frac{n-n_0}{N}) 2\pi i \frac{n-n_0}{N} - w_n \frac{1}{2} A_l^* \exp(-2\pi i \omega_l \frac{n-n_0}{N}) \left( -2\pi i \frac{n-n_0}{N} \right) \right)
 \end{aligned}$$

When applying

$$\frac{\partial f(x)g(x)}{\partial x} = g(x) \frac{\partial f(x)}{\partial x} + f(x) \frac{\partial g(x)}{\partial x}$$

and continuing the computation for each term separately, the first term yields

$$\begin{aligned}
& \sum_{n=0}^{N-1} 2 \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right] \\
& \quad \frac{\partial}{\partial \omega_p} \left( -w_n \frac{1}{2} A_l \exp(2\pi i \omega_l \frac{n-n_0}{N}) 2\pi i \frac{n-n_0}{N} - w_n \frac{1}{2} A_l^* \exp(-2\pi i \omega_l \frac{n-n_0}{N}) \left( -2\pi i \frac{n-n_0}{N} \right) \right) \\
= & \sum_{n=0}^{N-1} \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n-n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n-n_0}{N}) \right) \right] \\
& \quad \delta_{lp} \left( -w_n A_p \exp(2\pi i \omega_p \frac{n-n_0}{N}) \left( 2\pi i \frac{n-n_0}{N} \right)^2 \right. \\
& \quad \left. - w_n A_p^* \exp(-2\pi i \omega_p \frac{n-n_0}{N}) \left( -2\pi i \frac{n-n_0}{N} \right)^2 \right) \\
= & \delta_{lp} \sum_{n=0}^{N-1} \left[ \frac{1}{N} \sum_{m=0}^{N-1} R_m \exp(2\pi i m \frac{n-n_0}{N}) \right] \\
& \quad \left( -w_n A_p \exp(2\pi i \omega_p \frac{n-n_0}{N}) \left( 2\pi i \frac{n-n_0}{N} \right)^2 \right. \\
& \quad \left. - w_n A_p^* \exp(-2\pi i \omega_p \frac{n-n_0}{N}) \left( 2\pi i \frac{n-n_0}{N} \right)^2 \right) \\
= & -\delta_{lp} \frac{1}{N} \sum_{m=0}^{N-1} R_m \left[ A_p \sum_{n=0}^{N-1} w_n \left( 2\pi i \frac{n-n_0}{N} \right)^2 \exp(2\pi i(m+\omega_p) \frac{n-n_0}{N}) + \right. \\
& \quad \left. A_p^* \sum_{n=0}^{N-1} w_n \left( 2\pi i \frac{n-n_0}{N} \right)^2 \exp(2\pi i(m-\omega_p) \frac{n-n_0}{N}) \right] \\
= & -\delta_{lp} \frac{1}{N} \sum_{m=0}^{N-1} R_m \left[ A_p \sum_{n=0}^{N-1} w_n \left( 2\pi i \frac{n-n_0}{N} \right)^2 \exp(2\pi i(m+\omega_p) \frac{n-n_0}{N}) + \right. \\
& \quad \left. A_p^* \sum_{n=0}^{N-1} w_n \left( 2\pi i \frac{n-n_0}{N} \right)^2 \exp(2\pi i(m-\omega_p) \frac{n-n_0}{N}) \right] \\
= & -\delta_{lp} \frac{1}{N} \sum_{m=0}^{N-1} R_m (A_p W''(m+\omega_p) + A_p^* W''(m-\omega_p)) \\
= & -\delta_{lp} \frac{2}{N} \sum_{m=m_{min}}^{m_{max}} \Re(R_m A_p^* W''(m-\omega_p))
\end{aligned} \tag{40}$$

For the harmonic spectrum model one achieves

$$-\delta_{lp} \frac{2}{N} \sum_{q=1}^{Q_l-1} \sum_{m=m_{min}}^{m_{max}} \Re(R_m q^2 A_{p,q}^* W''(m-q\omega_p)) \tag{41}$$

where  $\delta_{lp}$  denotes the Kronecker symbol and the  $m_{min}$  and  $m_{max}$  values are dependent on  $q\omega_p$ .  $W''(m)$  can again be computed by one of the frequency response calculation methods

that were previously described. Note that,

$$\begin{aligned} W''(m) &= \frac{\partial W'(m)}{\partial m} \\ &= \frac{\partial}{\partial m} \sum_{n=0}^{N-1} w_n 2\pi i \frac{n - n_0}{N} \exp(2\pi i m \frac{n - n_0}{N}) \\ &= \sum_{n=0}^{N-1} w_n \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(2\pi i m \frac{n - n_0}{N}) \end{aligned}$$

5 Next, the second term is computed

$$\begin{aligned} &\sum_{n=0}^{N-1} \left( -w_n \frac{1}{2} A_l \exp(2\pi i \omega_l \frac{n - n_0}{N}) 2\pi i \frac{n - n_0}{N} - w_n \frac{1}{2} A_l^* \exp(-2\pi i \omega_l \frac{n - n_0}{N}) \left( -2\pi i \frac{n - n_0}{N} \right) \right) \\ &\quad \frac{\partial}{\partial \omega_p} 2 \left[ x_n - w_n \frac{1}{2} \sum_{k=0}^{K-1} \left( A_k \exp(2\pi i \omega_k \frac{n - n_0}{N}) + A_k^* \exp(-2\pi i \omega_k \frac{n - n_0}{N}) \right) \right] \\ &= \sum_{n=0}^{N-1} \left( -w_n \frac{1}{2} A_l \exp(2\pi i \omega_l \frac{n - n_0}{N}) 2\pi i \frac{n - n_0}{N} - w_n \frac{1}{2} A_l^* \exp(-2\pi i \omega_l \frac{n - n_0}{N}) \left( -2\pi i \frac{n - n_0}{N} \right) \right) \\ &\quad \left( -w_n A_p \exp(2\pi i \omega_p \frac{n - n_0}{N}) 2\pi i \frac{n - n_0}{N} - w_n A_p^* \exp(-2\pi i \omega_p \frac{n - n_0}{N}) \left( -2\pi i \frac{n - n_0}{N} \right) \right) \\ &= \frac{1}{2} \sum_{n=0}^{N-1} A_l A_p w_n^2 \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(2\pi i(\omega_l + \omega_p) \frac{n - n_0}{N}) - \\ &\quad \frac{1}{2} \sum_{n=0}^{N-1} A_l A_p^* w_n^2 \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(2\pi i(\omega_l - \omega_p) \frac{n - n_0}{N}) - \\ &\quad \frac{1}{2} \sum_{n=0}^{N-1} A_l^* A_p w_n^2 \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(-2\pi i(\omega_l - \omega_p) \frac{n - n_0}{N}) + \\ &\quad \frac{1}{2} \sum_{n=0}^{N-1} A_l^* A_p^* w_n^2 \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(-2\pi i(\omega_l + \omega_p) \frac{n - n_0}{N}) \\ &= \frac{1}{2} (A_l A_p Y''(\omega_l + \omega_p) - A_l A_p^* Y''(\omega_l - \omega_p) \\ &\quad - A_l^* A_p Y''(-\omega_l + \omega_p) + A_l^* A_p^* Y(-\omega_l - \omega_p)) \\ &= \Re(A_l A_p Y''(\omega_l + \omega_p) - \Re(A_l A_p^* Y''(\omega_l - \omega_p))) \end{aligned} \tag{42}$$

with

$$Y''(m) = \sum_{n=0}^{N-1} w_n^2 \left( 2\pi i \frac{n - n_0}{N} \right)^2 \exp(2\pi i m \frac{n - n_0}{N})$$

which is shown in Fig. 2. Here, a same conclusion can be drawn as for the matrix  $B$  which was used for the computation of the amplitudes. The first term of the hessian computation resulted only in nonzero values for the diagonal elements of the Hessian. The second term also produces significant values for the main diagonal elements since the frequency differences

fall in the main lobe of  $Y''(m)$ . Therefore, the Hessian is a band diagonal matrix, implying that only the elements close to the diagonal must be computed. In addition, the inverse matrix of the Hessian can be computed in linear time allowing to apply Newton's method.

For the harmonic model one achieves

$$\begin{aligned}
 & \frac{1}{2} \sum_{n=0}^{N-1} \left( w_n 2\pi i \frac{n - n_0}{N} \sum_{q=1}^{Q_p-1} q \left[ A_{p,q} \exp(2\pi i q \omega_p \frac{n - n_0}{N}) + A_{p,q}^* \exp(-2\pi i q \omega_p \frac{n - n_0}{N}) \right] \right) \\
 & \quad \left( w_n 2\pi i \frac{n - n_0}{N} \sum_{r=1}^{Q_l-1} r \left[ A_{l,r} \exp(2\pi i r \omega_l \frac{n - n_0}{N}) + A_{l,r}^* \exp(-2\pi i r \omega_l \frac{n - n_0}{N}) \right] \right) \\
 = & \frac{1}{2} \sum_{n=0}^{N-1} \left( w_n^2 (2\pi i \frac{n - n_0}{N})^2 \sum_{r=1}^{Q_l-1} \sum_{q=1}^{Q_p-1} qr \left[ A_{p,q} \exp(2\pi i q \omega_p \frac{n - n_0}{N}) + A_{p,q}^* \exp(-2\pi i q \omega_p \frac{n - n_0}{N}) \right. \right. \\
 & \quad \left. \left. \left[ A_{l,r} \exp(2\pi i r \omega_l \frac{n - n_0}{N}) + A_{l,r}^* \exp(-2\pi i r \omega_l \frac{n - n_0}{N}) \right] \right] \right) \\
 = & \frac{1}{2} \sum_{n=0}^{N-1} \left( w_n^2 (2\pi i \frac{n - n_0}{N})^2 \sum_{r=1}^{Q_l-1} \sum_{q=1}^{Q_p-1} qr \left[ A_{p,q} A_{l,r} \exp(2\pi i (q \omega_p + r \omega_l) \frac{n - n_0}{N}) \right. \right. \\
 & \quad A_{p,q} A_{l,r}^* \exp(2\pi i (q \omega_p - r \omega_l) \frac{n - n_0}{N}) + A_{p,q}^* A_{l,r} \exp(2\pi i (-q \omega_p + r \omega_l) \frac{n - n_0}{N}) + \\
 & \quad \left. \left. A_{p,q}^* A_{l,r}^* \exp(-2\pi i (q \omega_p + r \omega_l) \frac{n - n_0}{N}) \right] \right) \\
 = & \sum_{q=1}^{Q_p-1} \sum_{r=1}^{Q_l-1} qr [\Re(A_{p,q} A_{l,r} Y''(q \omega_p + r \omega_l)) - \Re(A_{p,q} A_{l,r}^* Y''(q \omega_p - r \omega_l))] \tag{43}
 \end{aligned}$$

Only components that have close frequencies must be evaluated due to the band limited nature of  $Y''$ . For instance for a given value  $q$ , and a given frequency response bandwidth  $\beta$ , only the  $r$  values must be considered for which  $r \omega_l$  falls in the main lobe. Since

$$\begin{aligned}
 0 \leq q \omega_p & \leq < \frac{N}{2} \\
 0 \leq r \omega_l & \leq < \frac{N}{2}
 \end{aligned}$$

the input values of  $Y''$  are bounded by

$$\begin{aligned}
 -\frac{N}{2} \leq q \omega_p - r \omega_l & \leq < \frac{N}{2} \\
 0 \leq q \omega_p + r \omega_l & \leq < N
 \end{aligned}$$

meaning that the main lobe of  $Y(q \omega_p - r \omega_l)$  ranges from  $-\frac{N}{M}\beta$  to  $\frac{N}{M}\beta$ . For  $Y(q \omega_p + r \omega_l)$  the main lobe is divided over the left and right side of the spectrum due to spectral replication yielding the intervals  $[0, \frac{N}{M}\beta]$  and  $[N - \frac{N}{M}\beta, N]$ . This implies that for  $Y(q \omega_p - r \omega_l)$  only the

$r$  values must be considered for which

$$\begin{aligned} -N\beta/M &\leq q\omega_p - r\omega_l \leq N\beta/M \\ \Rightarrow \frac{q\omega_p - N\beta/M}{\omega_l} &\leq r \leq \frac{q\omega_p + N\beta/M}{\omega_l} \end{aligned}$$

The two intervals for  $Y(q\omega_p + r\omega_l)$  yield

$$\begin{aligned} 5 \quad 0 &\leq q\omega_p + r\omega_l \leq N\beta/M \\ \Rightarrow \frac{-q\omega_p}{\omega_l} &\leq r \leq \frac{N\beta/M - q\omega_p}{\omega_l} \end{aligned}$$

and

$$\begin{aligned} N - N\beta/M &\leq q\omega_p + r\omega_l \leq N \\ \Rightarrow \frac{N(1 - \beta/M) - q\omega_p}{\omega_l} &\leq r \leq \frac{N - q\omega_p}{\omega_l} \end{aligned}$$

This results finally in

$$\begin{aligned} 10 \quad &\sum_{q=1}^{Q_p-1} \left[ \sum_{r=1}^{r_{max,1}} qr \Re(A_{p,q} A_{l,r} Y''(q\omega_p + r\omega_l)) \right. \\ &+ \sum_{r=r_{min,2}}^{r_{max,2}} qr \Re(A_{p,q} A_{l,r} Y''(q\omega_p + r\omega_l)) \\ &- \left. \sum_{r=r_{min,3}}^{r_{max,3}} qr \Re(A_{p,q} A_{l,r}^* Y''(q\omega_p - r\omega_l)) \right] \end{aligned}$$

with

$$\begin{aligned} r_{max,1} &= \lfloor \frac{N\beta/M - q\omega_p}{\omega_l} \rfloor \\ r_{min,2} &= \lceil \frac{N(1 - \beta/M) - q\omega_p}{\omega_l} \rceil \\ r_{max,2} &= \lfloor \frac{N - q\omega_p}{\omega_l} \rfloor \\ r_{min,3} &= \lceil \frac{q\omega_p - N\beta/M}{\omega_l} \rceil \\ r_{max,3} &= \lfloor \frac{q\omega_p + N\beta/M}{\omega_l} \rfloor \end{aligned}$$

15

For the harmonic model the hessian is not band diagonal but evidently, its size very small compared to the hessian for the non harmonic case. In addition, each element can be computed in linear time.

20 3.6 Efficient Frequency Optimisation

For the optimisation of the frequencies we distinguish a harmonic and a non-harmonic signal model. For the non-harmonic model, the Hessian matrix was shown to be band diagonal and invertible in linear time. Therefore only a shifted matrix of the Hessian denoted  $\overleftarrow{\mathbf{H}}$  needs to be computed. The frequencies are updated using

$$5 \quad \omega^{(r+1)} = \omega^{(r+1)} - \overleftarrow{\mathbf{H}}^{-1} \otimes h \quad (44)$$

where  $\otimes$  denotes the multiplication of a shifted matrix with a vector and  $(r)$  the index of the iteration. The outline of the computation is depicted in Figure 11.

For the harmonic model, the Hessian is not band diagonal but its very limited in size since normally only a few sources are available in the signal. The fundamental frequencies  
10 are optimized using

$$\omega^{(r+1)} = \omega^{(r)} - \mathbf{H}^{-1} h \quad (45)$$

The outline of the computation is depicted in Figure 12 and 13.

#### 4 The complete method

In figure 14, the complete analysis / synthesis method is depicted. This method can be  
15 applied for a harmonic or non-harmonic signal model.

First, the initial frequencies are computed (Figure 10). In the case of the harmonic model, a (multi)pitch estimator is used which determines an initial set of pitches from which a series of frequencies is computed for each source. For the non-harmonic model, peak picking can be used on the spectrum of the signal.

20 The preprocessing routine sorts the frequencies, removes frequencies that occur multiple times and determines how many diagonal bands  $D$  must be considered for the amplitude computation (Figure 9). Then, the amplitudes are computed (Figure 8). This is realized by computing the band diagonal elements of matrix  $\mathbf{B}$  and storing them in a shifted form  $\overleftarrow{\mathbf{B}}$ . The matrix  $\mathbf{C}$  is computed next and by solving the band limited system, the amplitudes are  
25 obtained.

The IFFT synthesizer computes the spectrum  $\tilde{X}_m$  from the amplitudes  $\tilde{A}$  and frequencies  $\bar{\omega}$ . The difference with the original spectrum  $X_m$  results in the residual spectrum  $R_m$  which is used for the frequency optimization in the next step.

In the case of the non-harmonic model, the Hessian matrix  $\mathbf{H}$  is band limited and is stored  
30 in a shifted form  $\overleftarrow{\mathbf{H}}$ . A second result of this property is that its inverse can be computed in linear time. The hessian and gradient are used to optimize the frequency values (Figure 11).

In the harmonic case, the fundamental frequencies of the different sources is optimized. Here, the Hessian is not band diagonal but it is very small since typically just a few sources are considered (Figure 12 and 13).

The iterative loop is continued until a stopping criterium is met. The results of the analysis  
5 are; the synthesized signal  $\bar{x}_n$ , the noise residual  $r_n$ , the amplitudes  $\bar{A}$  and frequencies  $\bar{\omega}$ . Note that the amplitudes are complex and therefore contain the phases.

## 5 Applications

The optimization of the computational efficiency of this highly accurate method significantly improves a large number of applications such as; multi-pitch extraction, parametric  
10 audio coding, source separation, audio classification, audio effects, automated transcription and annotation. These applications are depicted in Figure 15

### 5.1 High Resolution (Multi)Pitch Estimation

The efficient analysis method will improve pitch estimation techniques. For the current invention, the term "pitch" is used to denote the period of a harmonic source, and not the  
15 perceived height by the human auditory system. For instance, current (multi)pitch estimators based on autocorrelation such as the summary autocorrelation function (SACF) and the enhanced summary autocorrelation function (ESACF) allow to estimate multiple pitches. However, none of this methods takes into account the overlapping peaks that might occur. The frequency optimization for harmonic sources which is presented in this invention allows  
20 to improve the fundamental frequencies iteratively leading to very accurate pitch estimations. In addition, very small analysis windows can be used which enable to track fast variations in the pitch in an accurate manner.

### 5.2 Parametric Audio Coding

The resynthesis of the sound is of a very high quality indistinguishable from the original  
25 sound. In addition, the amplitudes and frequency parameters vary very slowly over time. Therefore, it is interesting to apply our method in the context of parametric coders where these parameters are stored in a differential manner what results in a considerable compression. Evidently this is interesting for the storage, transmission and broadcasting of digital audio.

### 5.3 Source Separation

When a multipitch estimator provides good initial values of the pitches the method optimizes all parameters so that an accurate match is obtained. By synthesizing each pitch component to a different signal, the different sound sources in the polyphonic recording can be separated.

#### 5.4 Automated Annotation and Transcription

Fast variations in the amplitude  $\bar{A}$  and frequencies  $\bar{\omega}$  indicate the beginning and end of a note. Therefore the method will contribute to the automatic annotation and/or transcription of the audio signal.

#### 5.5 Audio Effects

By modifying the frequencies and amplitudes of the different sinusoidal components high quality audio effects can be achieved. The power of this method lies in the fact that frequencies and amplitudes can be manipulated independently. This allows for instance time-stretching, sound morphing, pitch changes, timbre manipulation etc. all with a very high quality.

## DETAILED DESCRIPTION OF THE FIGURES

Figure 1 illustrates the band limited nature of the frequency response of the Blackmann-Harris window, denoted  $W(m)$ . Also the first and second derivative, denoted  $W'(m)$  and  $W''(m)$  respectively, are band diagonal.

5 Figure 2 illustrates frequency response of the zero padded Blackmann-Harris window  $W_M^N(m)$ , the squared Blackmann-Harris window  $Y(m)$  and its second derivative  $Y''(m)$ . Also these frequency responses are band limited.

Fig. 3 illustrates the theoretic motivation for a scaled table look-up. A time domain window  $w^N(n)$  1 is taken which is bandlimited in the frequency domain within a range 10  $[-\beta, \beta]$  4. When this window is zero padded up to a length  $P$  2 this results in a scaling in the frequency domain 5. Then, the spectrum is truncated in order to obtain a length  $N$  again 6. When taking the inverse fourier transform of this truncated spectrum, a window with length  $M$  zero padded up to a length  $N$  is obtained 3.

Figure 4 depicts a method, of computing the frequency response of a sampled zero padded window with a variable length, according to the invention, by a scaled look-up table. The look-up table 10 is pre-generated 7, 9, from a non-zero-padded analysis window  $w_n$  8. This results in a table 10 containing an oversampled frequency response  $W^N(m)$ . The frequency response of the zero-padded window with length  $M$  smaller than  $N$ , is generated in step 11 by doing a table look-up with the value  $m \frac{M}{N}$  returning the desired value 12.

20 Figure 5 depicts a method of computing the frequency response of a sampled zero padded window with a variable length, according to the invention, using an analytic frequency response calculator 13 as described in Eq. (14). For a given value  $m$  the calculator returns the frequency response of the zero padded window 14.

Figure 6 depicts a method of computing the frequency response of a sampled zero padded window with a variable length, according to the invention, using an interpolation calculator. 25 The frequency response of the zero padded sampled window is calculated using an interpolation method of the sampled frequency response 16 obtained by a FFT 15 of a zeros padded 20 window 19. The interpolator calculator 18 computes from the sampled frequency response 16 the variable length frequency response 17 using Eq. (16).

30 Figure 7 depicts the detail of a method of synthesizing a sound signal with variable length according to the invention. For each sinusoidal component  $k$  21, the range of values for  $m$  is determined which fall in the main lobe of the frequency response 22. For each sample 23 the

value of the frequency response with its center at  $\omega_k$  is computed 24 and multiplied with an amplitude  $A_k$  25 (Eq. 6). When this is repeated for each component  $k$ , the inverse fourier transform is taken 26 and the zero padding and imaginary part are removed 27. This results in the synthesize signal  $\tilde{x}_n$  28. A summary of this routine is denoted by 29. The synthesis has a complexity  $\mathcal{O}(N \log(N))$  since it contains an inverse FFT.

Figure 8 depicts the detail of a method of computing the amplitudes of the sinusoidal components in a sound signal in  $\mathcal{O}(N \log N)$  time, according to the invention. The amplitudes  $\bar{A}$  are computed 39 from a spectrum  $X_m$  for a given set of frequencies. This is realized by constructing the matrices  $C^1$ ,  $C^2$  34 and the matrices  $\overleftarrow{B}^{1,1}$ ,  $\overleftarrow{B}^{2,2}$  37 according to Eq. (25) and Eq. (30). By solving the set of equations represented by these matrices the amplitudes are computed 38. The matrices  $C^1$  and  $C^2$  are computed by determining for all partials  $l$  30 the range of  $m$  values 31, 32 of the main lobe and computing the value of each element according to Eq. (30) 34. The frequency response  $W_M^N$  is determined from the frequency response calculator of  $W_M^N(m)$ , 33. For the matrices  $B^{1,1}$  and  $B^{2,2}$ , only shifted matrices  $\overleftarrow{B}^{1,1}$  and  $\overleftarrow{B}^{2,2}$  are computed containing only its band diagonal elements. The width of the band is denoted  $D$ , For all  $k$  values from 0 to  $2D$  35 each row of the matrices  $\overleftarrow{B}^{1,1}$  and  $\overleftarrow{B}^{2,2}$  is computed 37 according to Eq. (25). The computation of these elements requires a variable lengths frequency response calculator  $Y_M^N(m)$ , 36. The equations denoted in Eq. (22) can now be solved directly on the shifted versions of  $B^{1,1}$ ,  $B^{2,2}$ , 38 yielding the amplitude values 39. A summary of the computation is denoted by 50

Figure 9 depicts the pre-processing that is needed before the computation of the amplitudes according to the invention. First the frequencies are sorted 40 in order to guarantee a band diagonal form of  $B$ . Each frequency 41, is compared with the next frequency, and when this difference is smaller than a value  $\epsilon$  42, the component is removed 43. This is done 25 to avoid the singularity of  $B$ . Then, the frequencies above and below  $\omega_l$  are considered 45 and the counter  $l$  44, 46 counts the number of components that fall in the main lobe with width  $\beta \frac{N}{M}$ . The maximum of this number  $D$  47, 48 over all components  $l$  yields the number of diagonals that must be taken in account. A short notation of the pre-processing is denoted by 51.

Figure 10 depicts several methods for the computation of the initial frequencies 56, 64, 62, according to the invention. A first method, takes the FFT 53 of the windowed signal  $x_n$  66 which is zero padded 52 up to a power 2 length. The local maxima are detected 55 from the spectrum  $X_m$  54 providing the initial frequency values. This method is well suited for the

non-harmonic signal model. A second method is more suited for the harmonic signal model. First a (multi)pitch estimator 58 is used such as the enhanced summary autocorrelation function (ESACF) which computes a number of pitches 59 from the windowed signal 57. For each detected pitch, a series of multiples of the pitch is produced 60 up to the Nyquist frequency. The series of frequencies of each source is denoted as  $\bar{\omega}_0, \bar{\omega}_1, \dots$ . These series together provide the initial values for the frequencies 62. Finally, an externally stored annotation or score 63 can be used which contains the pitches in the sound 65 or the individual frequency components.

Figure 11, depicts the frequency optimization for the non-harmonic model according to the embodiment of the invention. First the gradient and Hessian matrix are computed. Each element of the gradient  $h_l$ , 69, is computed with respect to Eq. (38) 72 over a range 70 of  $m$  values 71. In addition, a frequency response calculator is used for  $W_M^{IN}(m)$  74. Over the same range, the diagonal elements of the Hessian are computed according to Eq. (40) 72, using a frequency response calculator 74 for  $W_M^{IN}$ . Then the band diagonal elements of the Hessian are computed according to Eq. (42) 75 76 using a frequency response calculator for  $Y_M^{IN}$  77. The frequencies 79 are optimized by computing the inverse of the shifted Hessian matrix  $\tilde{H}$  and multiplying this with the gradient  $h$ , according to Eq. (44). Also short notation for the frequency optimization is given 138.

Figures 12 and 13 depict the frequency optimization for the harmonic model according to the embodiment of the invention. First, the gradient and Hessian matrix are computed. For each fundamental frequency  $\omega_l$ , 80, the gradient value  $h_l$  is computed over all partials  $q$  81. For each element of the gradient  $h_l$ , the range 82 of  $m$  values is determined 83 that lie in the main lobe of  $W_M^{IN}$ . The gradient is computed according to Eq. (39). This requires a frequency response calculator for  $W_M^{IN}(m)$  86. In the same loop, the diagonal elements of the Hessian are computed according to Eq. (40) 84, using a frequency response calculator 85 for  $W_M^{IN}$ . The method continues by going to the subroutine depicted in Figure 13 for the computation of the non-diagonal elements of the Hessian matrix 87, 91. Here, the elements  $H_{l,k}$  are computed with respect to Eq. (43) 95, 98, 102 using a frequency response calculator for  $Y_M^{IN}(m)$  99. The different ranges of  $r$  are determined 93, 94, 96, 97, 100, 101 that are required for this computation. Then the subroutine returns to the main routine 103, 88. Finally, the optimized frequencies are obtained 99 by computing Eq. (45) 89. Also here the short notation applies, 138.

Figure 14 depicts the complete Analysis/Synthesis method according to the embodiment

of the invention. Starting from a windowed short time signal  $x_n$  104, the initial values of the frequencies 107 are computed 105. These frequencies 112 are then pre-processed 108 and the number of diagonal bands  $D$  109 is determined. Then the amplitudes are computed 113 from the fourier transform 110, 111 of the signal  $X_m$ , the number of diagonal bands 109 and 5 the pre-processed frequencies 112. This results in the amplitudes 114 which together with the frequencies are used to synthesize the signal yielding  $\tilde{x}_n$  121  $\tilde{X}_m$  116. The difference 109 between the synthesized spectrum  $\tilde{X}_m$  116 and the original spectrum  $X_m$  111 yields the error spectrum  $R_m$  117. This error spectrum 117, the frequencies 112 and amplitudes 114 are used to optimize 106 the frequency values 106 for the next iteration. A stopping 10 criterium evaluator determines 118 whether the iteration is continued. Several criteria were described in section 3.3. When the criterium is met, the iteration is terminated 119. A short notation is depicted 122 which takes as input the signal  $x_n$  and produces a synthesized signal 123  $\tilde{x}_n$ , the noise residual  $r_n$ , the amplitudes  $\bar{A}$  and frequencies  $\bar{\omega}$ .

Figure (15) shows several applications of the analysis method according to the embodiment of the invention. In 139 the application of the analysis method as a high resolution pitch estimator is depicted. 140 depicts the use of the invention in the context of a parametric coder 124. At the sender side, the encoder converts the amplitudes  $\bar{A}$ , frequencies  $\omega$  and noise residual  $r_n$  to a bitstream 125 which can be stored, broadcasted or transmitted 126. At the receiver side, the decoder computes the amplitudes  $\bar{A}$ , frequencies  $\omega$  and noise 20 residual  $r_n$  back from the bitstream. The amplitudes and frequencies are then used for the synthesis of the deterministic part of the sound and by adding the noise, the signal is reconstructed. In 141 the use of the analysis method is depicted for high quality audio effects. An effects processor 130 produces the modified noise  $r_n^*$ , amplitudes  $\bar{A}^*$  and frequencies  $\bar{\omega}^*$  131 from which the processed sound signal 132, can be constructed 132. Fig. 142 shows 25 the application of the analysis method according to the invention to the separation of sound sources. All frequencies  $\bar{\omega}_i$  and amplitudes  $\bar{A}_i$  from each source are grouped 134 135, and then synthesize separately 136 resulting in the individual sources 137.

## CLAIMS

What we claim is

1. A method for modelling, i.a. analyzing and/or synthesizing, a windowed signal with a variable length, such as sound or speech signals, by computing the frequencies and complex amplitudes from the signal by using an optimized least squares method, whereby the method is executed using one or more variable length windows and has a complexity  $\mathcal{O}(N \log N)$ .
2. A method according to claim 1 using a harmonic signal model according to Eq. 32 and/or a non-harmonic signal model according to Eq. 1.
3. A method to compute the frequency response of a window with length  $M$  zero padded up to a length  $N$  comprising a scaled table look-up with a value  $m\frac{N}{M}$  resulting in the values  $W_M^N(m)$ .
4. A method to compute efficiently the frequency response of a window with length  $M$  zero padded up to a length  $N$  comprising the computation of the analytic frequency response as given by Eq. (12).
5. A method to compute efficiently the frequency response of a window with length  $M$  zero padded up to a length  $N$  consists of interpolating the sampled frequency response as given by Eq. (16).
6. A method according to any of claims 1 to 5 further comprising the step of:  
estimating the amplitudes using an optimized  $\mathcal{O}(N \log N)$  least squares estimation technique for variable length windows.
7. A method according to any of the previous claims 1 to 6, further comprising the step of:  
synthesizing a variable length sound signal by constructed its spectrum according to Eq. (6) and performing an inverse fourier transform whereby the synthesis has a time complexity of  $\mathcal{O}(N \log N)$ .
8. A method according to any of the previous claims 1 to 7 further comprising the step of:

optimizing the frequency values using Newton's method for the non-harmonic signal model according to Eq. (44) and/or the harmonic signal model according to Eq. (45).

9. A method according to any of the previous claims 1 to 8 further comprising:  
a stopping criterium evaluator.
10. a method according to any of the previous claims 1 to 9 further comprising the step of:  
determining the initial values of frequencies by peak picking from the spectrum.
11. A method according to any of the previous claims 1 to 10 further comprising the step of:  
determining the initial values of the fundamental frequencies for the harmonic signal model using preferably a multi-pitch estimator.
12. A method according to any of the previous claims 1 to 11 further comprising the step of:  
pre-processing before the amplitude estimation of claim 5, whereby this pre-processing comprises the steps of: sorting the frequencies, eliminating multiple occurring frequencies and determining the number of relevant diagonal bands.
13. A method of claim 12 further comprising the steps of computing the amplitudes by solving the equation given in Eq. (22), using From Eq. (25) such that only the elements around the diagonal of  $\mathbf{B}$  are taken into account, whereby a shifted form  $\overleftarrow{\mathbf{B}}$  is computed containing only  $D$  diagonal bands of  $\mathbf{B}$  according to Eq. (28) and Eq. (25), whereby the computation of the Eq. (25) requires the computation of the frequency response of the square window denoted  $Y(m)$ , a matrix  $\mathbf{C}$  is computed using Eq. (30), requiring computation of the frequency response of the window denoted  $W(m)$  preferably using the method of claim 2,3 or 4, and solving equation solving Eq. (22) on  $\overleftarrow{\mathbf{B}}$  and  $\mathbf{C}$  (Eq. (29)).
14. A method of claim 8 for the non-harmonic signal model, further comprising the step of:  
optimizing the computation of the gradient  $h$  and Hessian matrix  $\mathbf{H}$ , said gradient is computed from the residual spectrum  $R_m$ , the amplitude  $A_l$  and the derivative of the frequency response of the window denoted  $W'(m)$  as given by Eq. (38), whereby Eq. (42) and Eq. (40) imply that only the diagonal band of the Hessian matrix contains

significant values, whereby the Hessian is stored in a shifted form  $\tilde{H}$ , the diagonal elements of the Hessian are computed from the residual spectrum  $R_m$ , the amplitudes  $A_l$  and the second derivative of the frequency response of the window denoted  $W''(m)$  as given by Eq. (40), the second term of the Hessian given by Eq. (42) requires the second derivative of the frequency response of the square window, whereby all frequency responses are computed preferably using one of the methods in claim 2,3 or 4; and the diagonal Hessian is inverted in linear time and used to optimize the frequencies according to Eq. (44).

15. A method according to claim 8, for the harmonic signal model, further comprising the step of:

optimizing the computation of the gradient  $h$  and Hessian matrix  $H$ , whereby the gradient is computed from the residual spectrum  $R_m$ , the amplitudes  $A_{l,q}$  and the derivative of the frequency response of the window denoted  $W'(m)$  as given by Eq. (39), Eq. (43) and Eq. (41) compute the Hessian matrix  $H$  whereby all frequency responses are computed using preferably one of the methods in claim 2,3 or 4.

16. Apparatus wherein the method according to any of the previous claims 1 to 15 is implemented.

17. Use of a method according to any of the claims 1 to 15 or an apparatus according to claim 16 for accurate pitch estimation.

18. Use of a method according to any of the claims 1 to 15 or an apparatus according to claim 16 for parametric audio coders, where the noise residual, amplitudes and frequencies are encoded in a bitstream whereby said bitstream is stored, broadcasted or transmitted at the sender side, the receiver decodes the bitstream back to the parameters and synthesizes the sound.

19. Use of a method according to any of the claims 1 to 15 or an apparatus according to claim 16 for audio effects where the noise  $r_n$ , whereby the amplitudes  $\bar{A}$  and frequencies  $\bar{\omega}$  are manipulated by an effects processor yielding  $r_n^*$ ,  $\bar{A}^*$  and  $\bar{\omega}^*$ .

20. Use of a method according to any of the claims 1 to 15 or an apparatus according to claim 16 for source separation, by separating the frequencies and amplitudes from the same source, and synthesizing monophonic sources from these parameters separately.

21. Use of a method according to any of the claims 1 to 15 or an apparatus according to claim 16 for automated annotation and transcription whereby the signal is segmented according to the values of the amplitudes and frequencies.

**Abstract**

The present invention relates to a method for modelling, i.a. analyzing and/or  
5 synthesizing, a windowed signal with a variable length, such as sound or speech signals,  
by computing the frequencies and complex amplitudes from the signal by using an  
optimized least squares method, whereby the method is using one or more variable length  
windows and has a complexity  $\mathcal{O}(N \log N)$ .

1/13

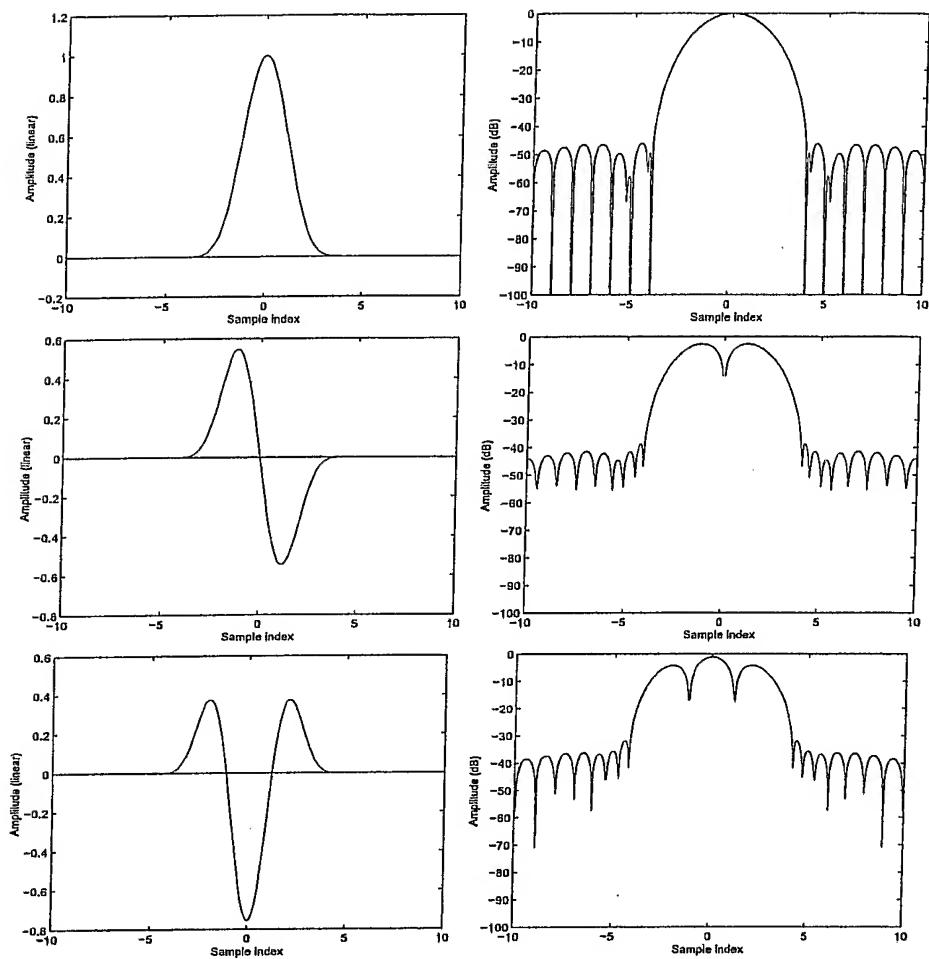


Figure 1: Top: Frequency Response of Blackmann-Harris  $W(m)$ , Middle: First Derivative  $W'(m)$ , Bottom: Second Derivative.  $W''(m)$

2/13

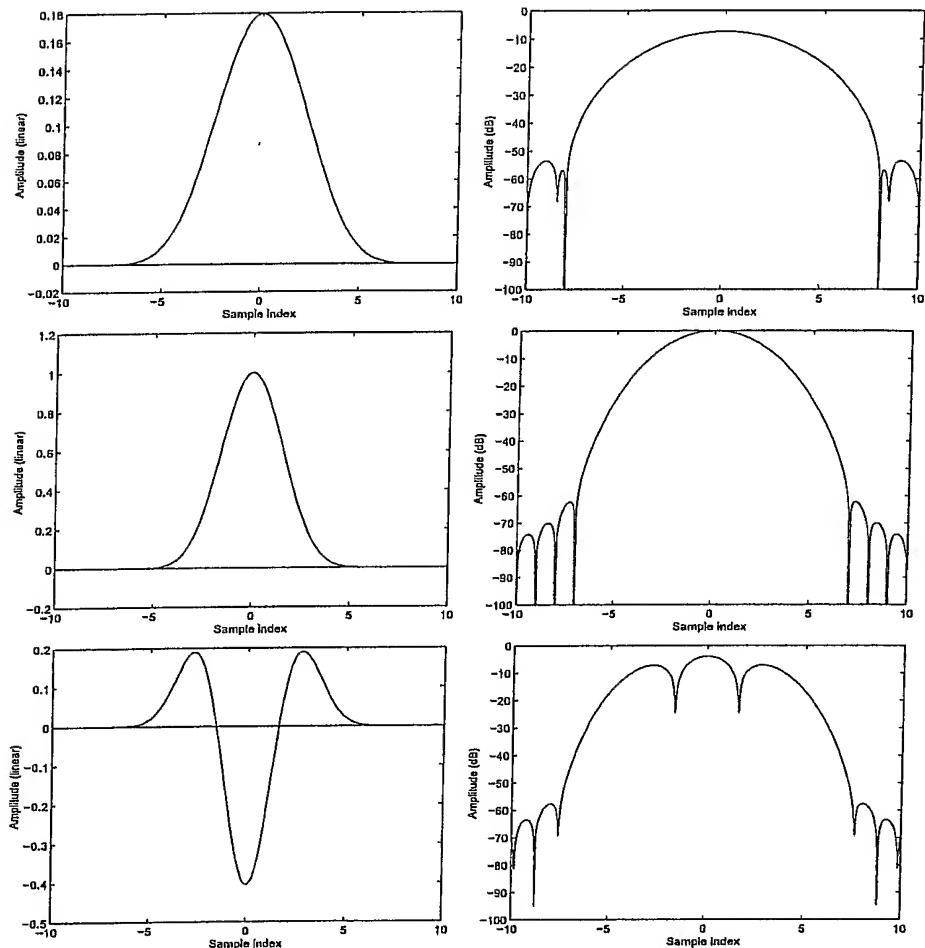


Figure 2: Top: Frequency Response of zero padded Blackmann-Harris Window  $\hat{W}_M^N(m)$ , Middle: Frequency Response of Squared Blackmann-Harris Window  $Y(m)$ , Bottom: Second Derivative of the Frequency Response of the Squared Blackmann-Harris Window  $Y''(m)$

3/13

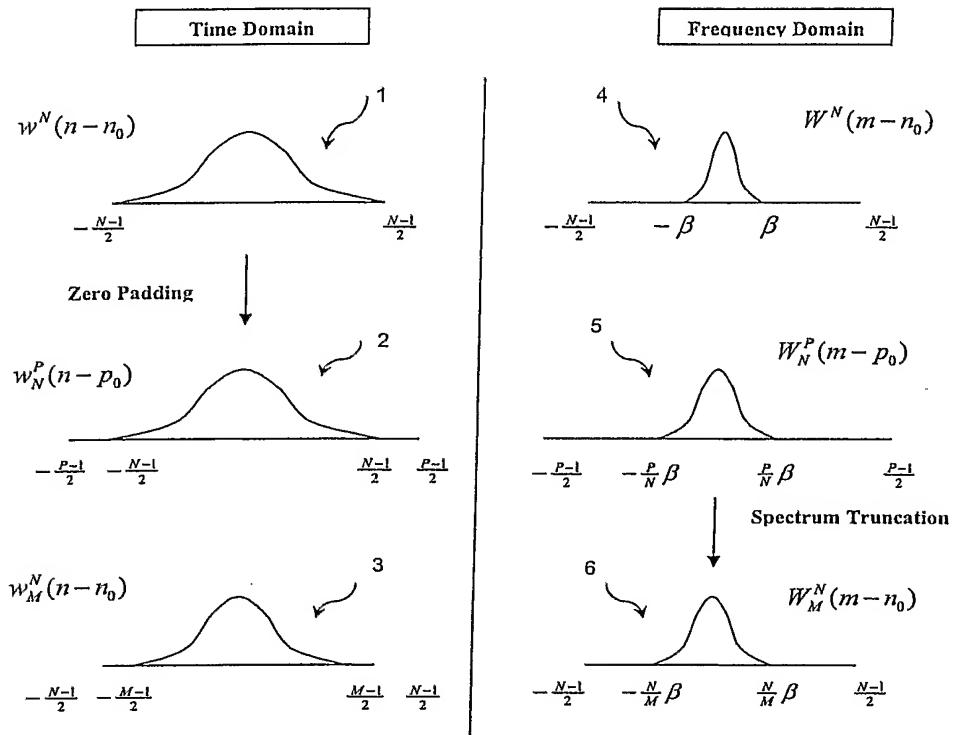


Figure 3: Theoretic Motivation for Scaled Table Look-up

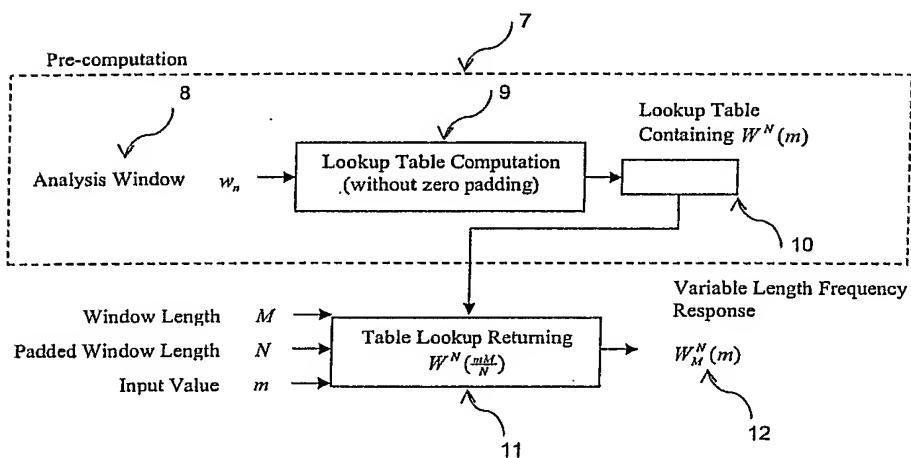


Figure 4: Scaled Table Look-up

4/13

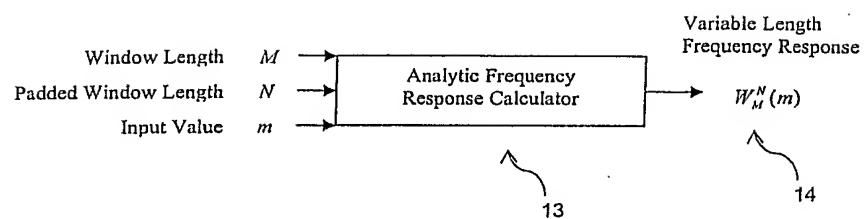


Figure 5: Analytic Frequency Response Computation

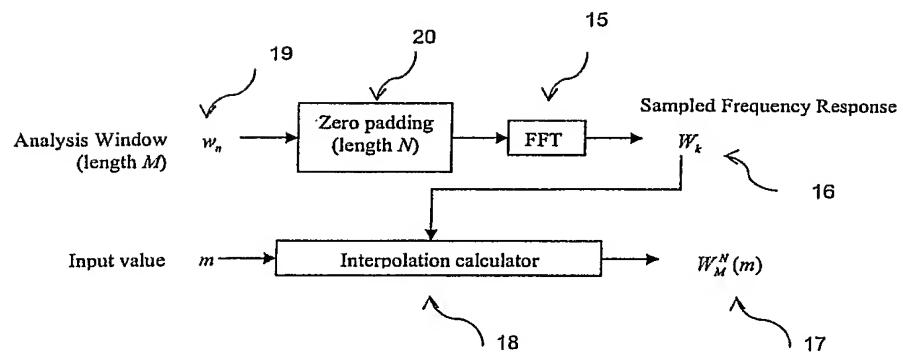


Figure 6: Sampled Frequency Response Interpolation

5/13

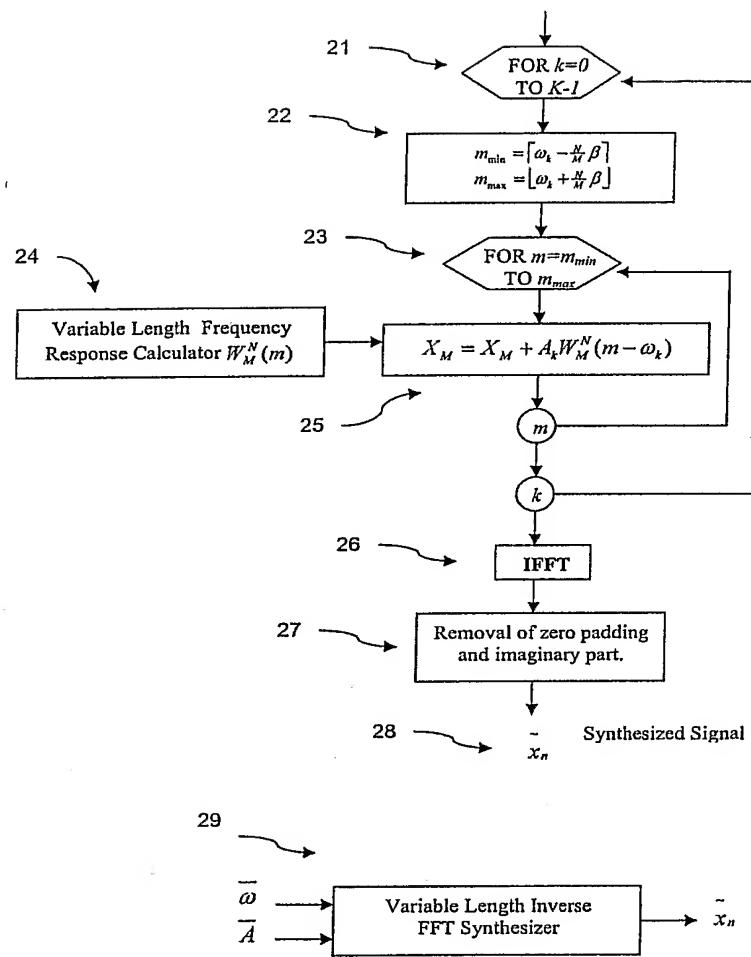


Figure 7: Synthesis

6/13

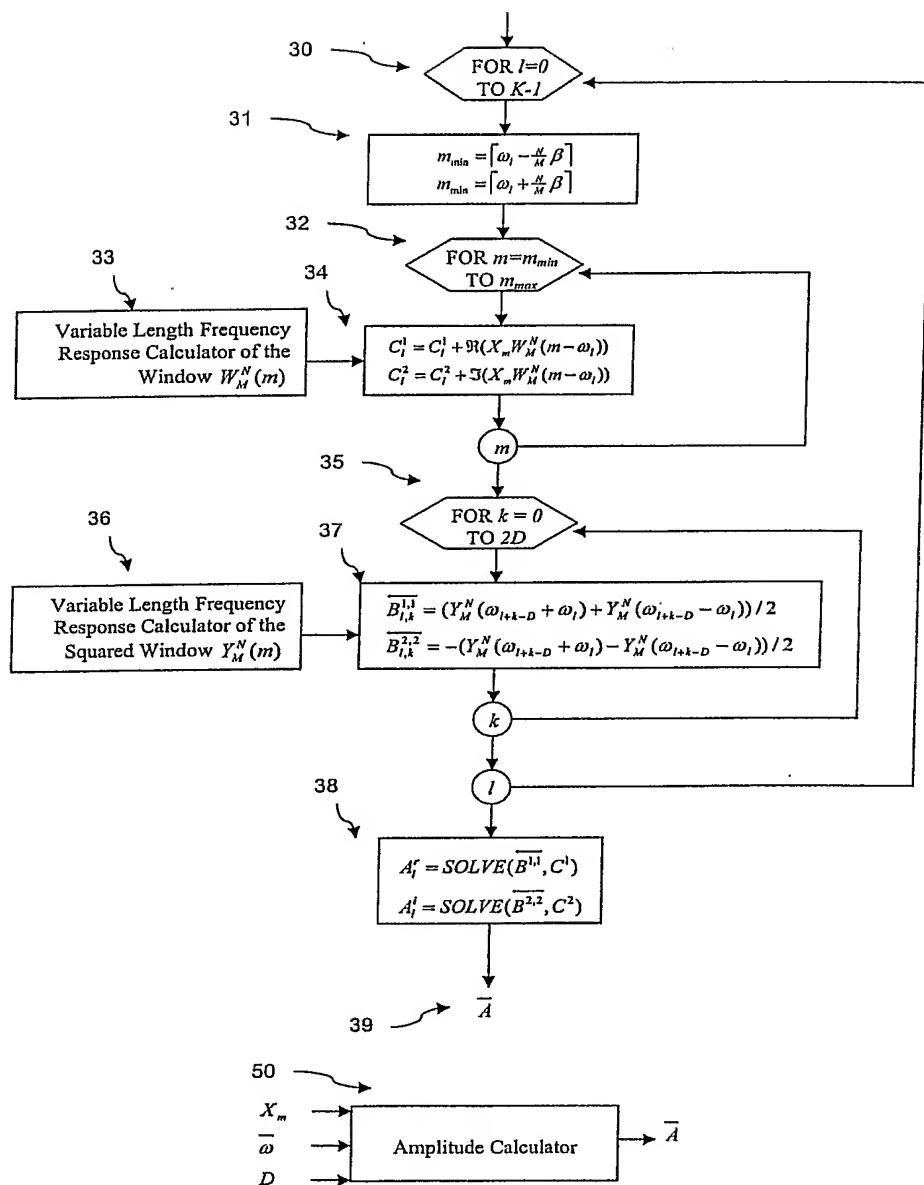


Figure 8: Amplitude Computation

7/13

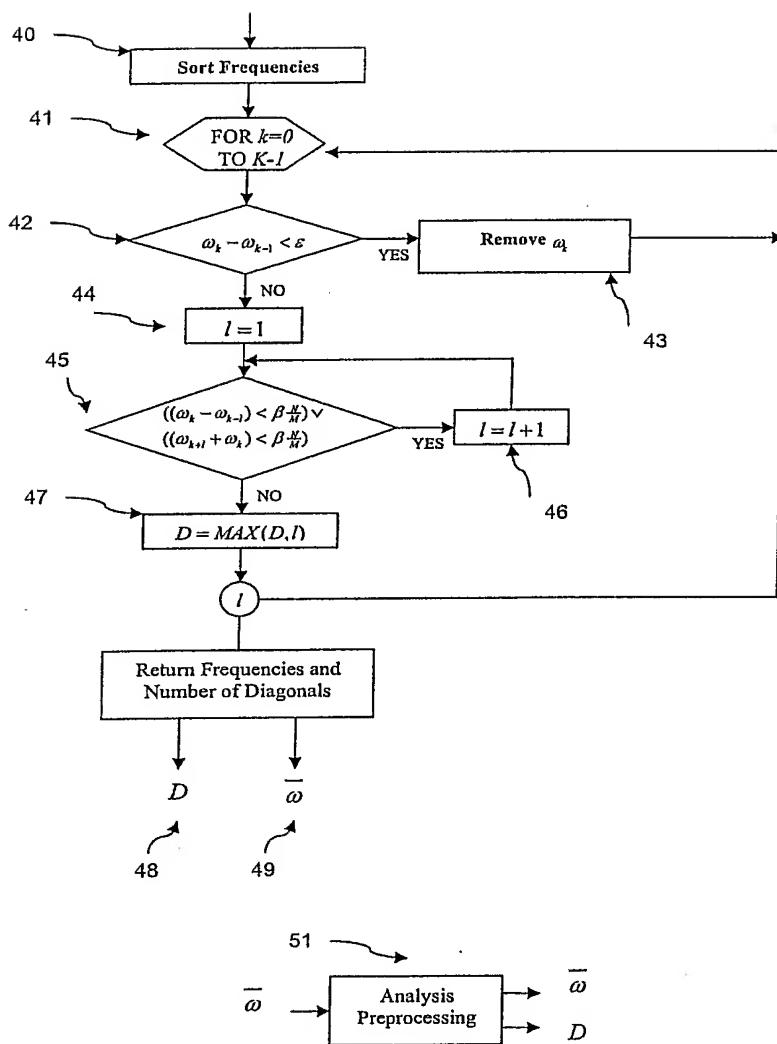


Figure 9: Preprocessing of the Amplitude Computation

8/13

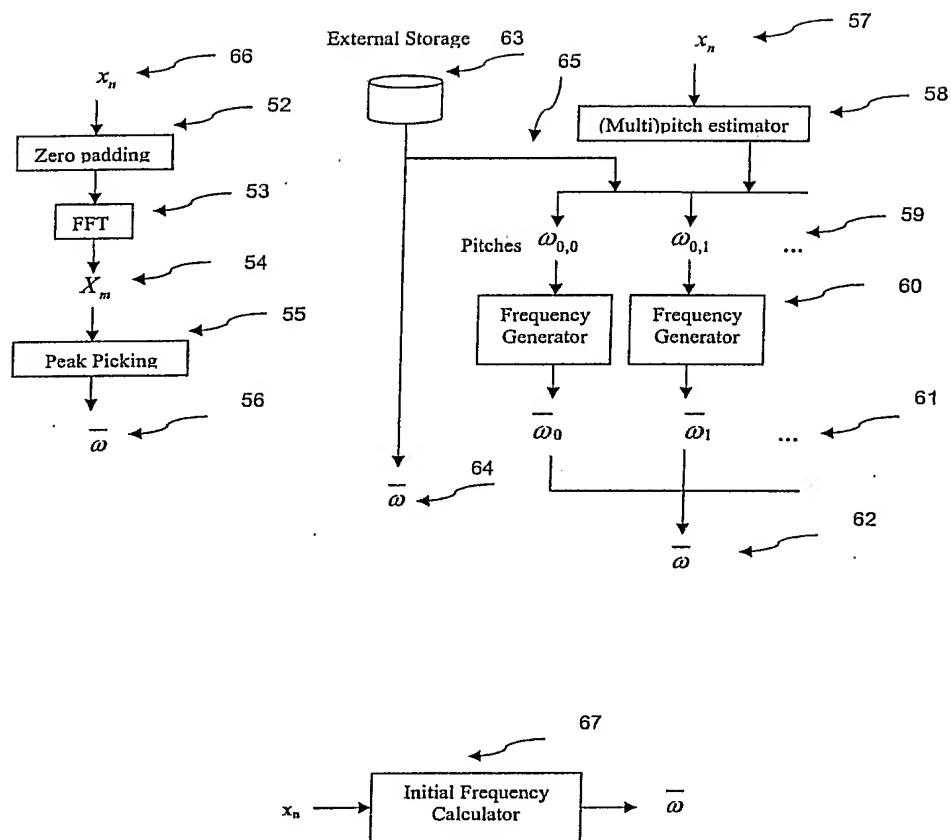


Figure 10: Initial Frequency Calculation

9 / 13

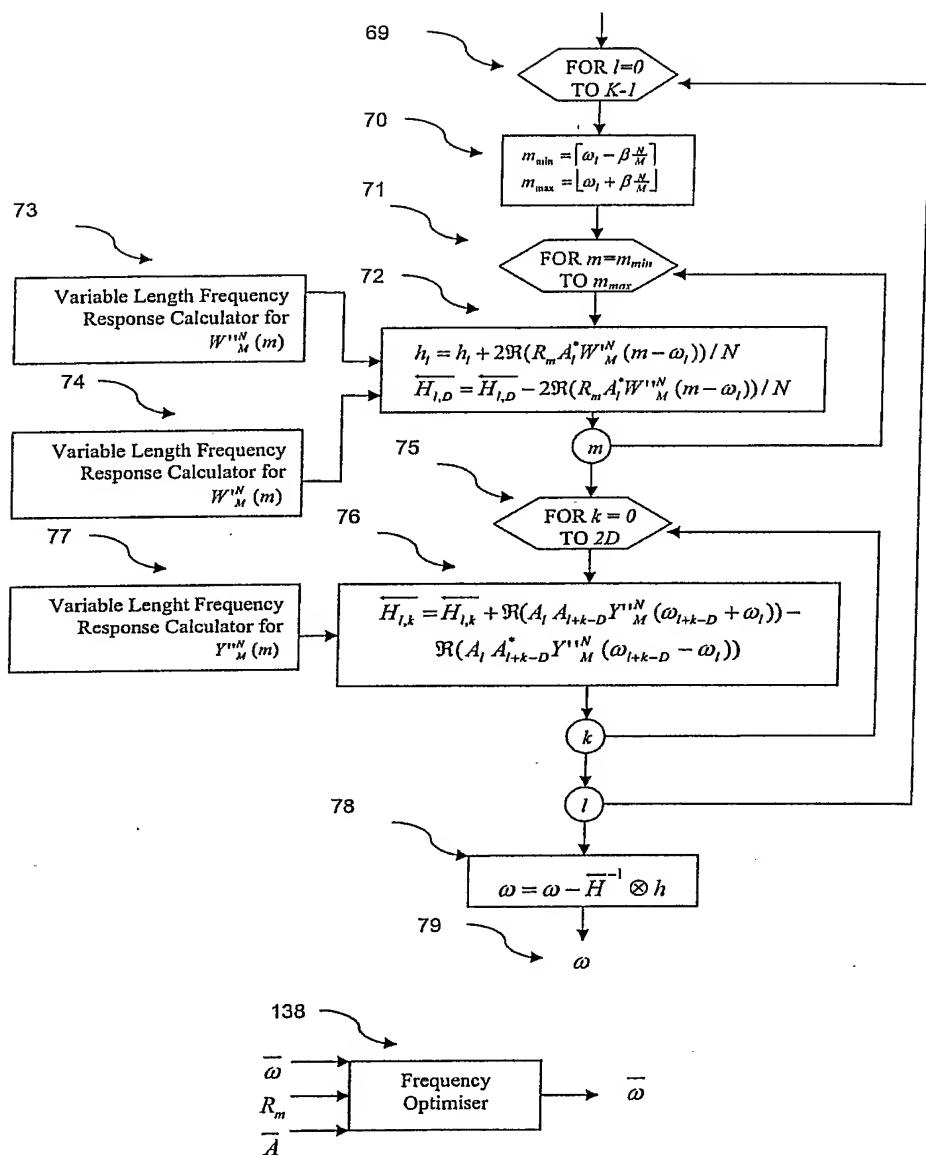


Figure 11: Frequency Optimiser for Non-Harmonic Model

10/13

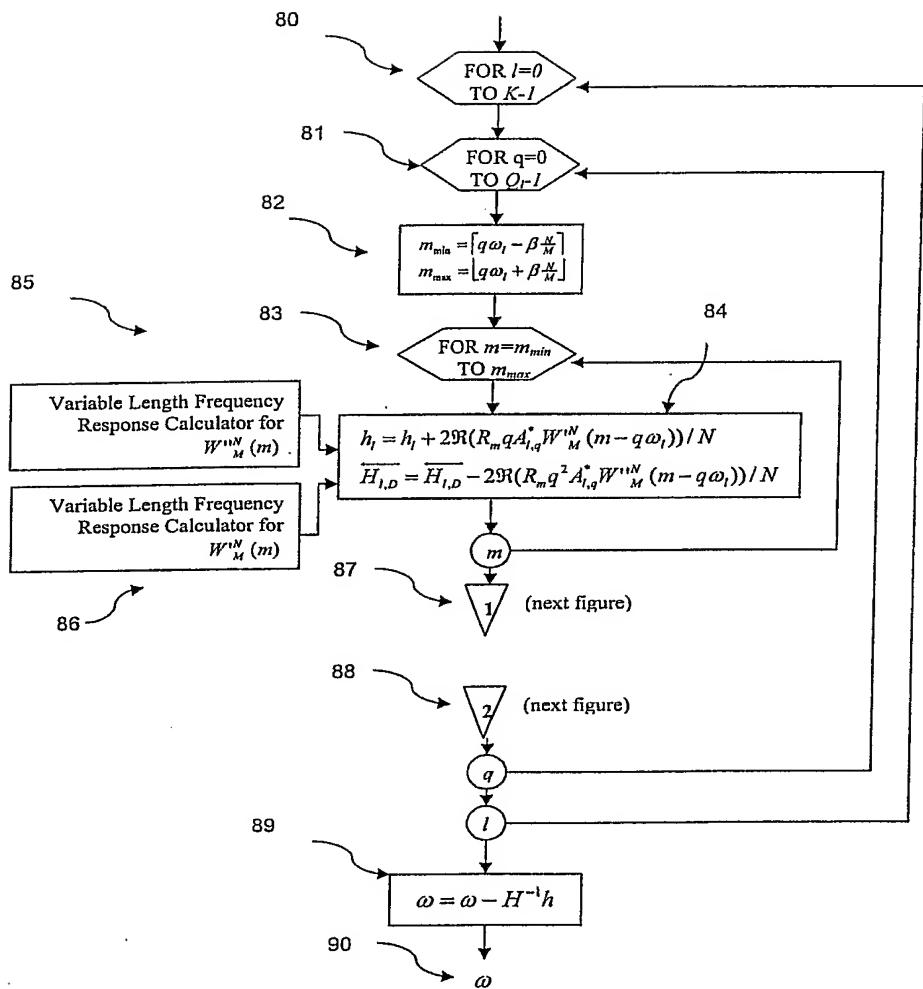


Figure 12: Frequency Optimiser for Harmonic Model

11/13

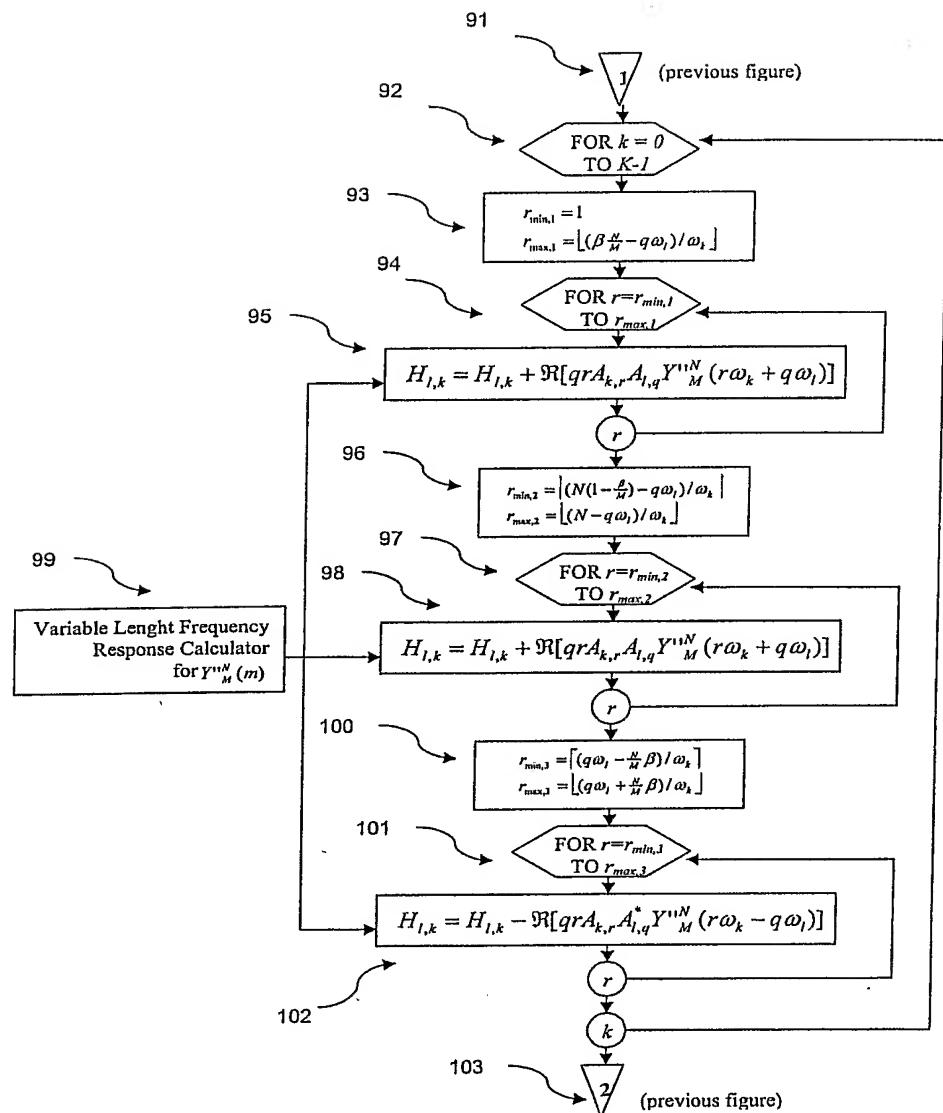


Figure 13: Frequency Optimiser for Harmonic Model (subroutine)

12/13

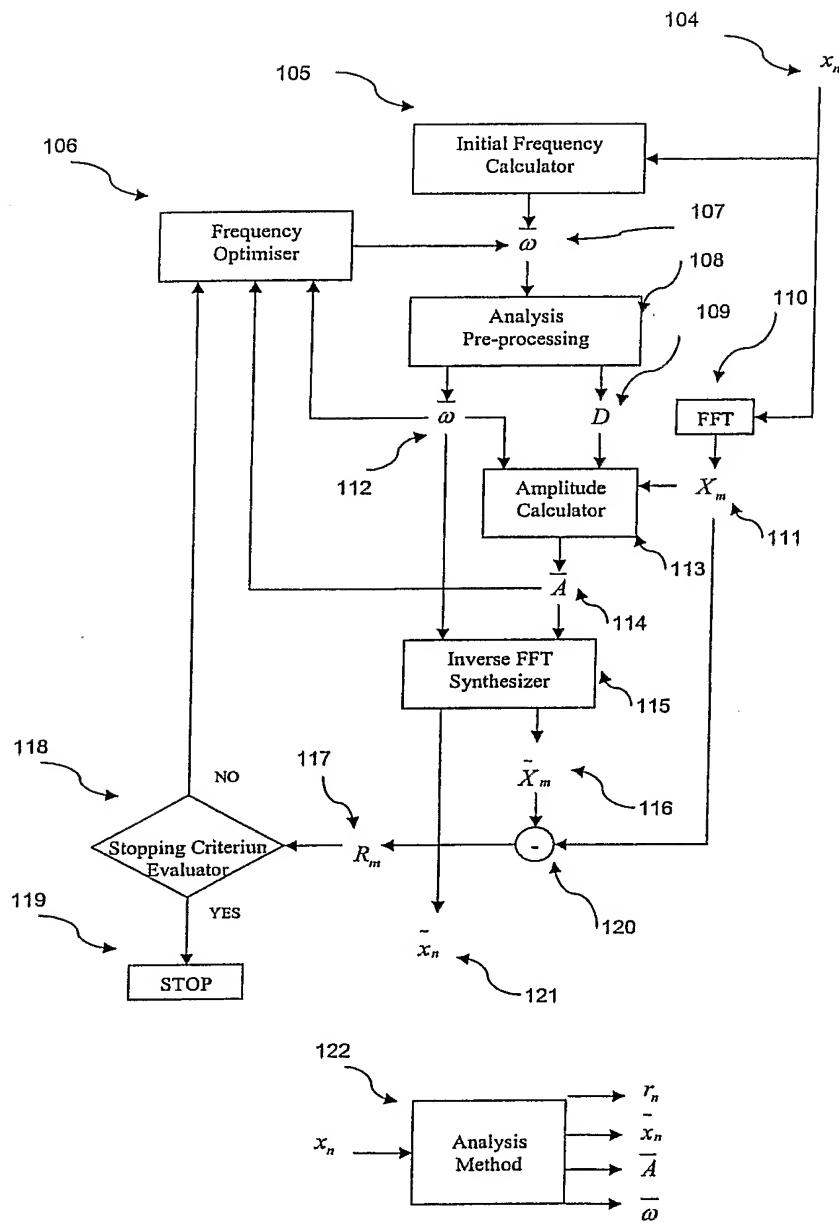


Figure 14: Complete Analysis/Synthesis Method

13/13

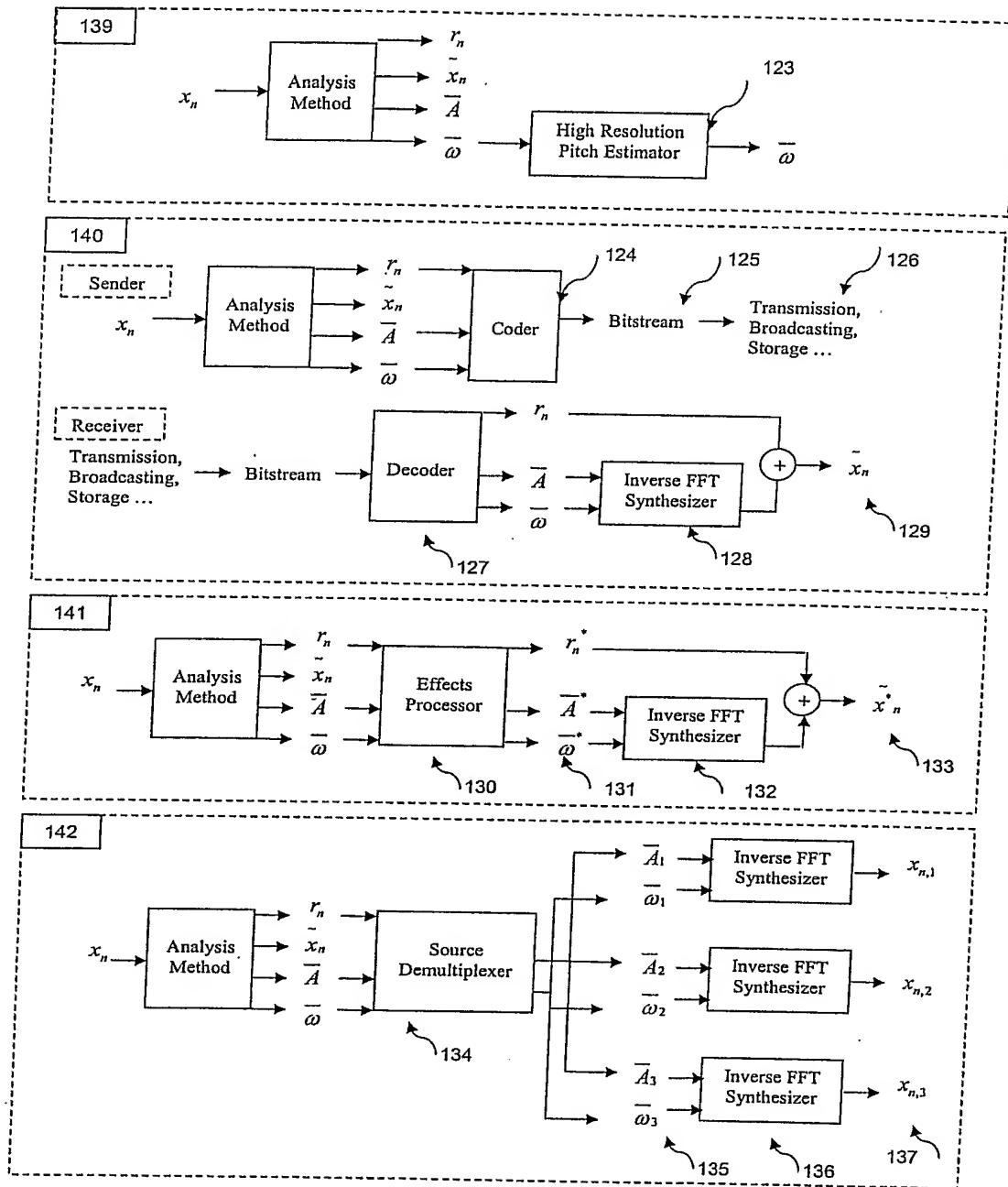


Figure 15: Applications; High Accuracy Pitch Estimation, Coding, Effects and Source Separation

